



**TUGAS AKHIR - SS141501**

**ANALISIS KLASIFIKASI PELANGGAN LISTRIK  
RUMAH TANGGA BERSUBSIDI KOTA SURABAYA  
MENGUNAKAN METODE *SUPPORT VECTOR  
MACHINE* DAN *NAÏVE BAYES CLASSIFIER***

**WIWIN YULI WIDIAWATI  
NRP 062116 4500 0015**

**Dosen Pembimbing  
Raden Mohamad Atok, Ph.D**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**



**TUGAS AKHIR - SS141501**

**ANALISIS KLASIFIKASI PELANGGAN LISTRIK  
RUMAH TANGGA BERSUBSIDI KOTA SURABAYA  
MENGUNAKAN METODE *SUPPORT VECTOR  
MACHINE* DAN *NAÏVE BAYES CLASSIFIER***

**WIWIN YULI WIDIAWATI  
NRP 062116 4500 0015**

**Dosen Pembimbing  
Raden Mohamad Atok, Ph.D**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**



**FINAL PROJECT - SS141501**

**CLASSIFICATION ANALYSIS OF SUBSIDIAL  
HOUSEHOLD ELECTRICAL CUSTOMERS IN  
SURABAYA USING SUPPORT VECTOR MACHINE  
AND NAÏVE BAYES CLASSIFIER**

**WIWIN YULI WIDIAWATI  
SN 062116 4500 0015**

**Supervisor  
Raden Mohamad Atok, Ph.D**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**



## LEMBAR PENGESAHAN

### ANALISIS KLASIFIKASI PELANGGAN LISTRIK RUMAH TANGGA BERSUBSIDI KOTA SURABAYA MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN *NAÏVE BAYES CLASSIFIER*

#### TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Sains  
pada

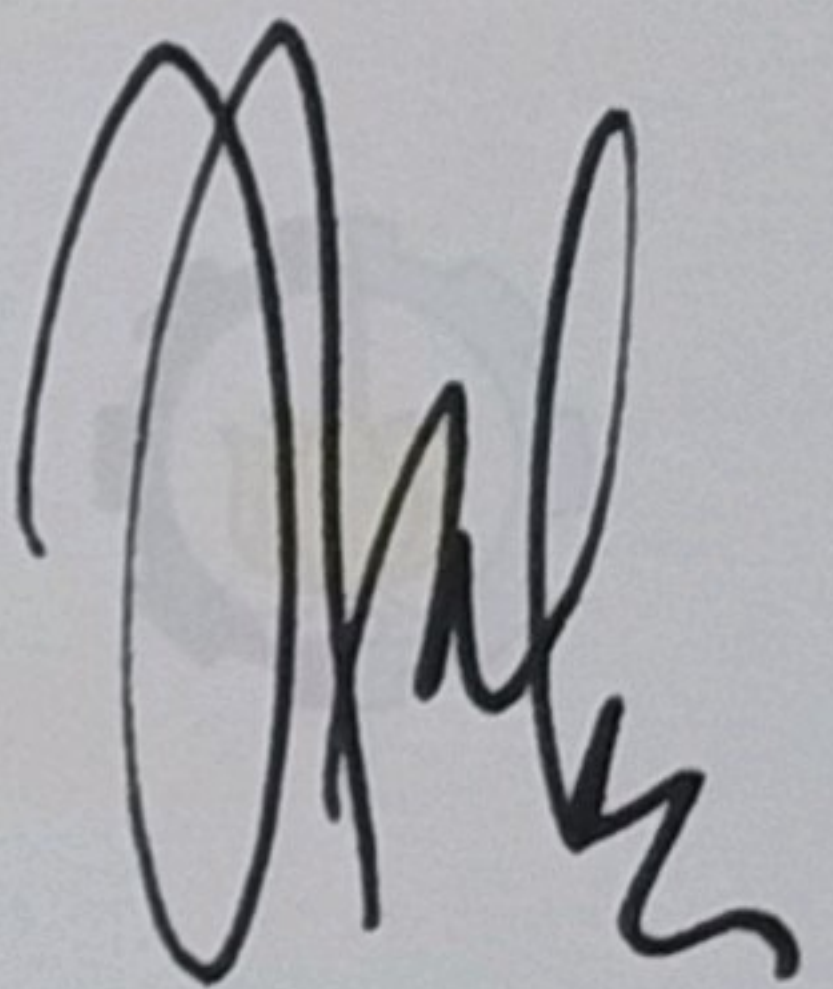
Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :

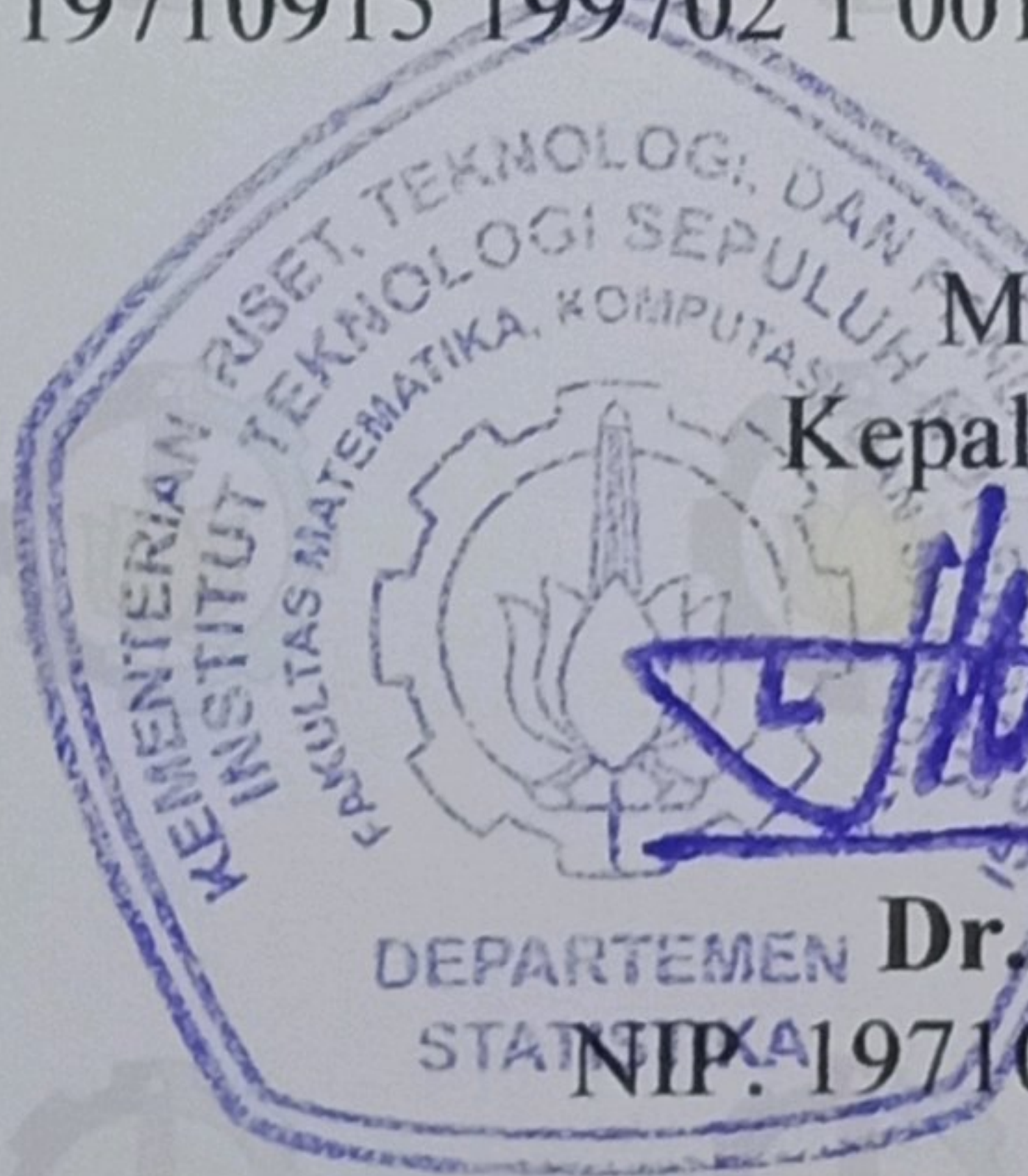
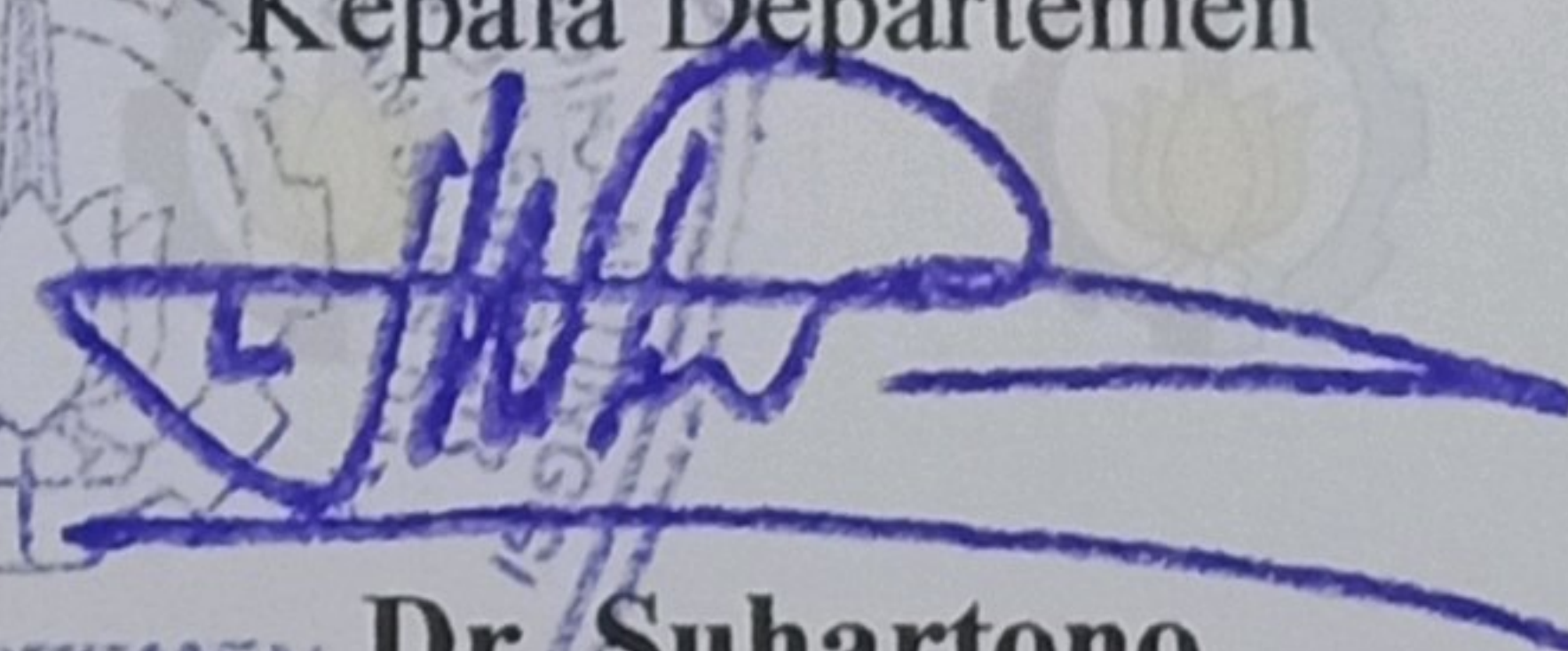
**Wiwin Yuli Widiawati**  
NRP. 062116 4500 0015

Disetujui oleh Pembimbing:

**Raden Mohamad Atok, Ph.D**  
NIP. 19710915 199702 1 001

(  )

Mengetahui,  
Kepala Departemen

  
  
**Dr. Suhartono**  
NIP. 19710929 199512 1 001

SURABAYA, JULI 2018



# **ANALISIS KLASIFIKASI PELANGGAN LISTRIK RUMAH TANGGA BERSUBSIDI KOTA SURABAYA MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN *NAÏVE BAYES CLASSIFIER***

**Nama Mahasiswa** : Wiwin Yuli Widiawati  
**NRP** : 06211645000015  
**Departemen** : Statistika  
**Pembimbing** : Raden Mohamad Atok, P.hD

## **Abstrak**

*Pemerintah menyediakan dana bagi kelompok masyarakat miskin dan tidak mampu dalam bentuk subsidi terhadap tarif tenaga listrik. Subsidi tersebut diberikan kepada pelanggan rumah tangga dengan daya 450 VA dan sebagian pelanggan 900 VA. Berdasarkan data dari Badan Pusat Statistik (BPS) yang telah diolah oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), dikeluarkan daftar 40% masyarakat Indonesia yang berada dalam kondisi ekonomi terendah dan menjadi prioritas pemerintah dalam program subsidi listrik. Pada penelitian ini bertujuan untuk melakukan klasifikasi sesuai karakteristik yang dimiliki oleh pelanggan subsidi listrik rumah tangga daya 450 VA dan 900 VA menggunakan metode Support Vector Machine dan Naïve Bayes Classifier. Hasil Perbandingan antara metode SVM dan NBC didapatkan bahwa SVM memberikan hasil yang lebih baik dengan parameter optimal pada kernel RBF adalah  $C = 10$  dan  $\gamma = 1$ . Kategori 450 VA sebanyak 91,6% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 8,4% pelanggan diprediksi dalam kategori subsidi 900 VA. Pada kategori subsidi 900 VA sebanyak 81,9% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 18,1% pelanggan diprediksi dalam kategori 450 VA.*

**Kata Kunci:** *Naive Bayes Classifier, Radial Basis Function, Subsidi Listrik, Support Vector Machine*

*(halaman ini sengaja dikosongkan)*

# CLASSIFICATION ANALYSIS OF SUBSIDIAL HOUSEHOLD ELECTRICAL CUSTOMERS IN SURABAYA USING SUPPORT VECTOR MACHINE AND NAÏVE BAYES CLASSIFIER

**Name** : Wiwin Yuli Widiawati  
**SN** : 06211645000015  
**Department** : Statistics  
**Supervisor** : Raden Mohamad Atok, P.hD

## **Abstract**

*The Government gives the electric subsidy funds for the poor and underprivileged society. Such subsidies are given to household customers with electrical power 450 VA and most customers 900 VA. Based on data from Badan Pusat Statistik (BPS) that had been treated by the national for acceleration of poverty reduction (TNP2K), released a list of 40% Indonesia society who are in the lowest economic conditions and become a priority of the Government in the electric subsidy program. This study aims to make classification household electrical subsidies customers based on their characteristics using the method of Support Vector Machine and Naïve Bayes Classifier. The results of the comparison between SVM and NBC acquired that SVM gives better results with optimal parameters in RBF kernel is  $C = 10$  and  $\gamma = 1$ . The Customers with 450 VA category who are 91.6% classified correctly and 8.4% predicted in the category 900 VA. The Customers with 900 VA category who are 81.9% classified correctly and 18.1% household predicted in the category 900 VA.*

**Keyword:** *Naive Bayes Classifier, Radial Basis Function, Electrical Subsidies, Support Vector Machine*

*(This page intentionally left blank)*



## KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul

**“Analisis Klasifikasi Pelanggan Listrik Rumah Tangga Bersubsidi Kota Surabaya Menggunakan Metode *Support Vector Machine* dan *Naive Bayes Classifier*”.**

Selain itu, sholawat serta salam penulis sampaikan kepada Nabi Muhammad SAW.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini tidak terlepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala ketulusan hati, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Bapak Raden Mohamad Atok, Ph.D., selaku dosen pembimbing atas segala saran, kesabaran, dan waktu yang diberikan kepada penulis selama pengerjaan Tugas Akhir ini.
2. Bapak Dr. Setiawan dan Dr. M. Sjahid Akbar selaku dosen penguji atas kritik dan saran yang membangun dalam penyusunan Tugas Akhir ini.
3. Bapak Dr. Suhartono selaku Kepala Departemen Statistika ITS, dan Bapak Dr. Sutikno selaku Ketua Program Studi S1 Statistika ITS yang telah membantu memfasilitasi untuk kelancaran penyelesaian Tugas Akhir ini.
4. Bapak Dr. Muhammad Mashuri sebagai dosen wali penulis. Terima kasih atas bantuan dan bimbingan kepada penulis selama kuliah di Statistika ITS.
5. Seluruh Dosen dan Karyawan Statistika ITS atas ilmu dan pengalaman yang dibagikan kepada penulis.
6. Ayah, Bunda, Adik dan (Almh.) Nenek tercinta beserta keluarga besar, terima kasih atas segala nasehat, penyemat, dukungan, doa, serta kasih sayang yang tiada ternilai kepada penulis.
7. Teman-teman Lintas Jalur Statistika ITS Angkatan 2016 khususnya teman-teman Lintas Jalur kelas B, Bella, Muwa,

Annisa, Cicil, Mbak Ima, Mifta, Ima, Vio, Nanda, Evi, Elok, Ardi, Aldi, terimakasih atas semangat dan kenangan yang luar biasa selama 2 tahun ini.

8. Teman-teman Kos Gang Makam No. 3, khususnya kak Dina, mbak Rifqoh, Izzah, Via, terima kasih atas kenangan, semangat, dan dukungannya.
9. Teman-teman SPARTA IPA A SMAN 1 Pamekasan, terima kasih atas dukungan dan semangatnya.
10. Semua pihak yang tidak dapat disebutkan satu-persatu yang telah membantu hingga pelaksanaan Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa tidak ada kesempurnaan dalam penciptaan manusia, begitu pula dalam penulisan laporan Tugas Akhir ini. Oleh karena itu kritik dan saran dari semua pihak sangat diharapkan untuk perbaikan penelitian selanjutnya. Semoga Tugas Akhir ini bermanfaat dan memberikan wawasan bagi semua pihak.

Surabaya, Juli 2018

Penulis



# DAFTAR ISI

	Halaman
<b>JUDUL</b> .....	i
<b>LEMBAR PENGESAHAN</b> .....	iii
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vii
<b>KATA PENGANTAR</b> .....	ix
<b>DAFTAR ISI</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xiii
<b>DAFTAR TABEL</b> .....	xv
<b>DAFTAR LAMPIRAN</b> .....	xvii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan .....	5
1.4 Batasan Masalah .....	5
1.5 Manfaat .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
2.1 Klasifikasi .....	7
2.2 <i>K-Fold Cross Validation</i> .....	7
2.3 <i>Fast Correlation Based Filter (FCBF)</i> .....	8
2.4 <i>Support Vector Machine (SVM)</i> .....	9
2.5 <i>Naïve Bayes Classifier</i> .....	15
2.6 Pengukuran Akurasi Klasifikasi .....	17
2.7 Energi Lisrik Sektor Rumah Tangga .....	18
<b>BAB III METODOLOGI PENELITIAN</b> .....	19
3.1 Sumber Data .....	19
3.2 Variabel Penelitian.....	20
3.3 Langkah Analisis .....	22
3.4 Diagram Alir.....	25
<b>BAB IV ANALISIS DAN PEMBAHASAN</b> .....	27
4.1 Karakteristik Rumah Tangga Subsidi Listrik.....	27
4.2 Klasifikasi <i>Support Vector Machine</i> .....	34
4.3 Klasifikasi <i>Naïve Bayes Classifier</i> .....	42

4.4 Perbandingan Performa Klasifikasi.....	47
BAB V KESIMPULAN DAN SARAN .....	53
5.1 Kesimpulan .....	53
5.2 Saran .....	53
DAFTAR PUSTAKA.....	55
LAMPIRAN .....	59



## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b>	Ilustrasi Pembagian Data .....8
<b>Gambar 2.2</b>	Konsep <i>Hyperplane</i> SVM Linier .....10
<b>Gambar 2.3</b>	<i>Hyperplane</i> SVM Linier .....11
<b>Gambar 2.4</b>	<i>Hyperplane</i> Non Linier .....13
<b>Gambar 3.1</b>	Diagram Alir Penelitian .....25
<b>Gambar 4.1</b>	Pengaruh Nilai Cost dan Gamma pada Hasil Akurasi.....37
<b>Gambar 4.2</b>	Pengaruh Nilai Cost dan Gamma pada Hasil AUC .....38
<b>Gambar 4.3</b>	Pengaruh Nilai Cost dan Gamma pada Hasil <i>G-Mean</i> .....38

*(Halaman ini sengaja dikosongkan)*



## DAFTAR TABEL

	Halaman
<b>Tabel 2.1</b> <i>Confusion Matrix</i> .....	17
<b>Tabel 2.2</b> Jumlah Pelanggan Rumah Tangga .....	18
<b>Tabel 3.1</b> Struktur Data Penelitian.....	19
<b>Tabel 3.2</b> Variabel Penelitian.....	20
<b>Tabel 3.3</b> Variabel Penelitian (Lanjutan).....	21
<b>Tabel 4.1</b> Karakteristik Jumlah Anggota Keluarga .....	27
<b>Tabel 4.2</b> Karakteristik Status Kepemilikan Bangunan Tempat Tinggal .....	28
<b>Tabel 4.3</b> Karakteristik Luas Lantai.....	28
<b>Tabel 4.4</b> Karakteristik Jenis Lantai Terluas .....	29
<b>Tabel 4.5</b> Karakteristik Jenis Dinding Terluas .....	29
<b>Tabel 4.6</b> Karakteristik Jenis Atap Terluas.....	29
<b>Tabel 4.7</b> Karakteristik Jumlah Kamar Tidur .....	30
<b>Tabel 4.8</b> Karakteristik Sumber Air Minum.....	30
<b>Tabel 4.9</b> Karakteristik Bahan Bakar Memasak.....	31
<b>Tabel 4.10</b> Karakteristik Kepemilikan Lemari Es .....	31
<b>Tabel 4.11</b> Kepemilikan Televisi .....	31
<b>Tabel 4.12</b> Kepemilikan Komputer atau Laptop .....	32
<b>Tabel 4.13</b> Kepemilikan Sepeda .....	32
<b>Tabel 4.14</b> Kepemilikan Sepeda Motor .....	32
<b>Tabel 4.15</b> Karakteristik Peserta Program Raskin.....	33
<b>Tabel 4.16</b> Karakteristik Status Kesejahteraan.....	33
<b>Tabel 4.17</b> Hasil Seleksi Variabel.....	34
<b>Tabel 4.18</b> Hasil Klasifikasi Kernel Linier ( <i>Training</i> ).....	35
<b>Tabel 4.19</b> Hasil Klasifikasi Kernel Linier ( <i>Testing</i> ) .....	36
<b>Tabel 4.20</b> Hasil Klasifikasi Kernel RBF ( <i>Training</i> ).....	37
<b>Tabel 4.21</b> Hasil Klasifikasi Kernel RBF ( <i>Testing</i> ) .....	39
<b>Tabel 4.22</b> Hasil FCBF Klasifikasi Kernel Linier ( <i>Training</i> )....	40
<b>Tabel 4.23</b> Hasil FCBF Klasifikasi Kernel Linier ( <i>Testing</i> ) .....	40
<b>Tabel 4.24</b> Hasil FCBF Klasifikasi Kernel RBF ( <i>Training</i> ) .....	41
<b>Tabel 4.25</b> Hasil FCBF Klasifikasi Kernel RBF ( <i>Testing</i> ) .....	42
<b>Tabel 4.26</b> Hasil Klasifikasi <i>Naïve Bayes Classifier</i> .....	42

<b>Tabel 4.27</b>	Hasil FCBF Klasifikasi <i>Naïve Bayes Classifier</i> .....	43
<b>Tabel 4.28</b>	Peluang Prior .....	43
<b>Tabel 4.29</b>	Peluang Kategori Variabel Prediktor .....	44
<b>Tabel 4.30</b>	Peluang Kategori Variabel Prediktor (Lanjutan).....	45
<b>Tabel 4.31</b>	Peluang Kategori Variabel Prediktor (Lanjutan).....	46
<b>Tabel 4.32</b>	Peluang Parsial Data <i>Testing</i> Pertama .....	46
<b>Tabel 4.33</b>	Perbandingan Performa Klasifikasi .....	47
<b>Tabel 4.34</b>	Hasil <i>Confusion Matrix</i> .....	48
<b>Tabel 4.35</b>	Hasil Klasifikasi Jumlah Kamar Tidur .....	49
<b>Tabel 4.36</b>	Hasil Klasifikasi Kepemilikan Lemari Es .....	49
<b>Tabel 4.37</b>	Hasil Klasifikasi Kepemilikan Sepeda .....	50
<b>Tabel 4.38</b>	Hasil Klasifikasi Status Kesejahteraan .....	50

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1.</b> <i>Syntax Metode Support Vector Machine</i> .....	59
<b>Lampiran 2.</b> <i>Syntax Metode Naïve Bayes Classifier</i> .....	59
<b>Lampiran 3.</b> <i>Output Support Vector Machine Kernel RBF</i> ....	61
<b>Lampiran 4.</b> <i>Output Support Vector Machine Kernel</i> <i>RBF Seleksi Varibel FCBF</i> .....	66
<b>Lampiran 5.</b> <i>Output Support Vector Machine</i> <i>Kernel Linier</i> .....	71
<b>Lampiran 6.</b> <i>Output Support Vector Machine Kernel</i> <i>Linier Seleksi Varibel FCBF</i> .....	73
<b>Lampiran 7.</b> <i>Output Naïve Bayes Classifier</i> .....	75



*(Halaman ini sengaja dikosongkan)*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Listrik merupakan energi vital bagi keberlangsungan aktivitas manusia baik bagi individu, kelompok masyarakat maupun dunia industri. Energi listrik dapat dimanfaatkan untuk melakukan aktivitas dengan manfaat yang sangat besar dimana berbagai peralatan untuk memenuhi kebutuhan hidup dioperasikan dengan menggunakan energi listrik (Siregar & Warman, 2013). Kegiatan masyarakat yang cenderung meningkat dari waktu ke waktu, mendorong peningkatan pengoperasian peralatan dengan tenaga listrik. PT. Perusahaan Listrik Negara (Persero) yang selanjutnya disebut sebagai PLN merupakan perusahaan penyedia energi listrik dengan mengelompokkan pelanggan menjadi 5 golongan berdasarkan keperluan pemakaian (Saleh, 2015). Golongan tersebut antara lain sektor rumah tangga, industri, bisnis, sosial, dan penerangan umum. Golongan rumah tangga merupakan kelompok pemakai energi listrik paling besar yaitu mencapai 90% dari total pelanggan (PLN, 2017). Jumlah rumah tangga yang ada di Provinsi Jawa Timur mengalami kenaikan setiap tahunnya, sehingga permintaan terhadap sambungan listrik disektor tersebut juga mengalami peningkatan. Hal ini juga tidak terlepas karena Jawa Timur memiliki jumlah penduduk terpadat kedua di Indonesia (BPS, 2017). PT PLN (Persero) memiliki wilayah kerja di seluruh Indonesia, salah satunya adalah PT PLN (Persero) Distribusi Jawa Timur yang menargetkan rasio elektrifikasi di provinsi Jawa Timur dapat mencapai 100% pada tahun 2019 dari posisi saat ini 91%.

Pemerintah mencanangkan program 35.000 MW dan Program Indonesia Terang, khususnya di wilayah Indonesia bagian Timur dalam bentuk pembangunan infrastruktur penyediaan tenaga listrik. Sesuai dengan Undang-Undang Nomor 30 Tahun 2009 tentang ketenagalistrikan bahwa untuk penyediaan tenaga listrik, Pemerintah menyediakan dana bagi kelompok

masyarakat tidak mampu, pembangunan sarana penyediaan tenaga listrik di daerah yang belum berkembang, pembangunan tenaga listrik di daerah terpencil dan perbatasan, serta pembangunan listrik di perdesaan. Penyediaan dana bagi kelompok masyarakat miskin dan tidak mampu diberikan dalam bentuk subsidi terhadap tarif tenaga listrik untuk pelanggan rumah tangga dengan daya 450 VA dan 900 VA. Subsidi listrik adalah bantuan yang diberikan Pemerintah kepada konsumen dalam bentuk Tarif Tenaga Listrik (TTL) yang lebih rendah dibandingkan dengan tarif keekonomiannya, dengan jumlah pemakaian listrik yang sama, konsumen yang memperoleh tarif bersubsidi akan membayar rekening atau tagihan listrik lebih rendah daripada konsumen yang tidak mendapatkan subsidi. Besaran subsidi listrik yang diterima konsumen rumah tangga daya 450 VA dan 900 VA tergantung pada jumlah pemakaian energi listriknya. Secara rata-rata, konsumen rumah tangga daya 450 VA mendapatkan subsidi listrik sebesar Rp 80.000 per konsumen per bulan, dan untuk konsumen rumah tangga daya 900 VA adalah rata-rata Rp 90.000 per konsumen per bulan (PLN, 2017). Pembangunan subsidi listrik tersebut membutuhkan anggaran yang sangat besar. Pemerintah perlu memastikan bahwa penerima manfaat subsidi listrik adalah hanya masyarakat miskin dan tidak mampu. Penghematan subsidi listrik akan digunakan untuk membangun sarana penyediaan tenaga listrik dalam rangka mengakselerasi rasio elektrifikasi.

Pemerintah telah menetapkan kebijakan program subsidi listrik tepat sasaran pada tanggal 1 Januari 2017 (ESDM, 2016). Bagi masyarakat yang tidak mampu secara ekonomi dan belum terdaftar sebagai penerima bantuan program listrik subsidi, dibuka kesempatan untuk melaporkan diri melalui kecamatan, sehingga dapat diberikan arahan untuk mendaftar diri sebagai calon penerima program subsidi listrik. Berdasarkan data dari Badan Pusat Statistik (BPS) yang telah diolah oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), dikeluarkan daftar 40% masyarakat Indonesia yang berada dalam kondisi



ekonomi terendah (ESDM, 2016). Data yang dikeluarkan oleh TNP2K tersebut menjadi database rumah tangga yang menjadi prioritas dalam program penanggulangan kemiskinan di Indonesia. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan klasifikasi sesuai karakteristik yang dimiliki oleh pelanggan subsidi listrik rumah tangga daya 450 VA dan 900 VA dengan menerapkan metode data mining. PLN merupakan salah satu perusahaan yang bisa memanfaatkan metode data mining karena permintaan sambungan listrik oleh pelanggan yang berlangsung setiap hari sehingga semakin lama data pelanggan yang tersimpan sangatlah besar.

Data mining merupakan suatu proses untuk mengidentifikasi pola yang memiliki potensi dan berguna untuk mengelola dataset yang besar (Witten, I. H., Frank, E., & Hall, 2011). Dalam data mining terdapat 10 algoritma teratas yang paling berpengaruh yang dipilih oleh peneliti dalam komunitas data mining, dimana enam diantaranya adalah algoritma klasifikasi yaitu *Support Vector Machines* (SVM), *AdaBoost*, *K-Nearest Neighbor* (K-NN), *Naive Bayes Classifier* (NBC) dan CART (Wu & Kumar, 2009). Salah satu algoritma yang banyak diteliti adalah *Support Vector Machine* (SVM). SVM merupakan suatu teknik untuk menemukan *hyperplane* yang bisa memisahkan dua set data dari dua kelas yang berbeda (Cortes & Vapnik, 1995). SVM dikenal sebagai teknik *machine learning* yang dapat digunakan untuk menemukan fungsi pemisah (*classifier*) yang optimal. Metode ini digunakan karena mampu melakukan klasifikasi pada data yang bisa dipisahkan oleh suatu garis linier maupun nonlinier). SVM merupakan salah satu metode terbaik yang dapat digunakan dalam masalah klasifikasi pola serta kemampuan generalisasi, yaitu kemampuan untuk mengklasifikasikan data lain yang tidak termasuk dalam data yang dipakai pada *machine learning* (Gun, 1998). Selanjutnya juga akan dilakukan klasifikasi menggunakan metode *Naive Bayes Classifier* (NBC) sebagai perbandingan terhadap metode SVM. NBC adalah salah satu klasifikasi probabilistik paling sederhana yang didasarkan pada teorema

Bayes dan mencari peluang terbesar dari kemungkinan klasifikasi. NBC dapat melakukan proses klasifikasi dengan lebih cepat dibandingkan dengan metode lain yang lebih kompleks (Zhang, 2004). Meskipun termasuk dalam klasifikasi yang sederhana, namun dalam beberapa penelitian menunjukkan bahwa metode NBC memiliki performa yang baik dibandingkan metode lain yang lebih kompleks. Penelitian yang dilakukan oleh Nayak dan Natarajan (2016) memberikan kesimpulan bahwa metode NBC memiliki nilai akurasi yang lebih tinggi dibandingkan SVM dan *Random Forest*.

## 1.2 Rumusan Masalah

Penyediaan dana bagi kelompok masyarakat miskin dan tidak mampu diberikan dalam bentuk subsidi terhadap tarif tenaga listrik untuk pelanggan rumah tangga dengan daya listrik 450 VA dan 900 VA. Pemerintah perlu memastikan bahwa penerima manfaat subsidi listrik adalah hanya masyarakat miskin dan tidak mampu. Berdasarkan data dari Badan Pusat Statistik (BPS) yang telah diolah oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), dikeluarkan daftar 40% masyarakat Indonesia yang berada dalam kondisi ekonomi terendah. Data yang dikeluarkan oleh TNP2K tersebut menjadi database rumah tangga yang menjadi prioritas dalam program penanggulangan kemiskinan di Indonesia. Permasalahan yang dikaji dalam penelitian ini adalah bagaimana melakukan klasifikasi sesuai karakteristik yang dimiliki oleh pelanggan subsidi listrik rumah tangga daya 450 VA dan 900 VA dengan menggunakan metode *Support Vector Machine* dan *Naïve Bayes Classifier*. SVM merupakan salah satu metode terbaik yang dapat digunakan dalam masalah klasifikasi pola, serta kemampuan generalisasi, yaitu kemampuan untuk mengklasifikasikan data lain yang tidak termasuk dalam data yang dipakai pada *machine learning*. Sedangkan *Naïve Bayes Classifier* dapat melakukan proses klasifikasi dengan lebih cepat serta memiliki performa yang baik dibandingkan metode lain yang lebih kompleks.

### 1.3 Tujuan

Tujuan dari penelitian ini berdasarkan rumusan masalah adalah sebagai berikut.

1. Melakukan klasifikasi sesuai karakteristik yang dimiliki oleh pelanggan subsidi listrik rumah tangga daya 450 VA dan 900 VA menggunakan metode *Support Vector Machine* dan *Naïve Bayes Classifier*.
2. Membandingkan ketepatan klasifikasi antara metode *Support Vector Machine* dan *Naïve Bayes Classifier*.
3. Melakukan analisis karakteristik pelanggan subsidi listrik daya 450 VA dan 900 VA sesuai hasil klasifikasi.

### 1.4 Batasan Masalah

Penelitian ini dilakukan pada pelanggan rumah tangga Wilayah Kota Surabaya yang memperoleh subsidi listrik daya 450 VA dan 900 VA (pelanggan subsidi). Data yang diperoleh dari PT PLN yaitu data pelanggan rumah tangga bulan Januari 2018 yang memperoleh subsidi listrik, dan dilakukan penyesuaian dengan data fakir miskin dari Basis Data Terpadu (BDT) program perlindungan sosial yang dikelola oleh Badan Perencanaan Kota Surabaya. Penyesuaian data PLN dan Bappeko tersebut hanya berdasarkan data nama kepala rumah tangga pelanggan yang memperoleh subsidi listrik.

### 1.5 Manfaat

Hasil dari penelitian ini diharapkan dapat menjadi alternatif model untuk menentukan daya yang sesuai karakteristik rumah tangga agar memudahkan dalam mengenal pelanggan yang layak menerima subsidi listrik dari pemerintah agar tepat sasaran. Selain itu, penelitian ini dapat memberikan informasi keakuratan klasifikasi dalam penerapan metode *Support Vector Machine* dan *Naïve Bayes Classifier*.



*(Halaman ini sengaja dikosongkan)*

## **BAB II**

### **TINJAUAN PUSTAKA**

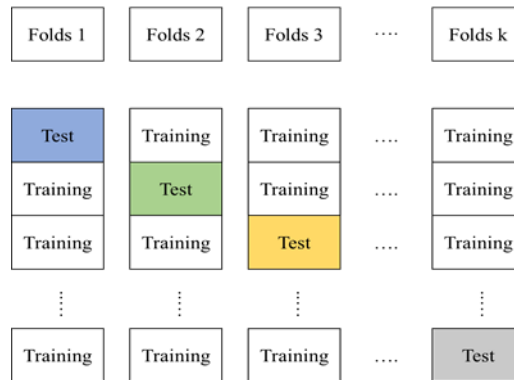
#### **2.1 Klasifikasi**

Klasifikasi adalah proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang kelasnya tidak diketahui (Han & Kamber, 2006). Selain itu, klasifikasi merupakan fungsi pembelajaran yang memetakan (mengklasifikasi) sebuah unsur data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan. Klasifikasi merupakan bagian dari data mining, dimana data mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining juga proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *classification trees*, *naïve bayes classifiers*, *neural networks*, *rough sets*, *k-nearest neighbor*, *rule based method*, *memory based reasoning*, dan *support vector machines*. Klasifikasi data terdiri dari dua langkah proses yang dilakukan, pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk aturan klasifikasi yaitu untuk membangun model atau fungsi. Proses kedua adalah klasifikasi, dimana data *testing* digunakan untuk memperkirakan akurasi dari hasil klasifikasi (Han & Kamber, 2006).

#### **2.2 K-Fold Cross Validation**

*K-fold cross validation* adalah suatu metode yang digunakan untuk mempartisi data menjadi data *training* dan data *testing*. sebuah teknik yang menggunakan keseluruhan *dataset* yang ada sebagai *training dan testing*. Metode ini banyak digunakan peneliti karena dapat mengurangi bias yang terjadi pada saat

pengambilan sampel. *K-fold cross validation* secara berulang-ulang membagi data menjadi data *training* dan data *testing*, dimana setiap data mendapat kesempatan menjadi data *testing* (Gokgoz & Subasi, 2015). Metode ini ditujukan agar diperoleh tingkat kepercayaan yang tinggi karena semua *dataset* dilibatkan sebagai data *training* maupun *testing* (Bengio, 2004).



**Gambar 2.1** Ilustrasi Pembagian Data

### 2.3 Fast Correlation Based Filter (FCBF)

FCBF merupakan salah satu algoritma seleksi variabel yang bersifat multivariat, mengukur kelas dan korelasi antar variabel. Secara umum, suatu variabel dikatakan bagus jika variabel tersebut relevan dengan konsep kelas namun tidak redundan pada variabel yang lain. Jika diterapkan korelasi antara dua variabel sebagai ukuran kebaikan, maka sebuah variabel dikatakan bagus untuk klasifikasi jika berkorelasi sangat tinggi dengan kelas namun tidak berkorelasi dengan variabel lainnya. Namun pengukuran dengan korelasi tidak mampu menangkap korelasi yang tidak *linear* selain itu korelasi mengharuskan semua variabel dan kelas mengandung nilai numerik. Untuk mengatasi kekurangan ini, Yu dan Liu (2009), menerapkan pendekatan lain yaitu memilih ukuran korelasi berdasarkan konsep *information theoretical entropy*. *Entropy* dari variabel X didefinisikan pada persamaan (2.1).

$$H(X) = -\sum_{i=1}^n P(x_i) \log_2(P(x_i)) \quad (2.1)$$

*Entropy* dari variabel X jika diketahui variabel Y didefinisikan pada persamaan (2.2).

$$H(X|Y) = -\sum_{i=1}^n P(y_i) \sum_{i=1}^n P(x_i|y_i) \log_2(P(x_i|y_i)) \quad (2.2)$$

$P(x_i)$  adalah posterior probabilities untuk semua nilai X dan  $P(x_i|y_i)$  adalah posterior probabilities dari X jika Y diketahui. Berdasarkan *entropy* tersebut dapat diperoleh *Information Gain* pada persamaan (2.3).

$$IG(X|Y) = H(X) - H(X|Y). \quad (2.3)$$

Korelasi antar variabel dapat diukur menggunakan *symmetrical uncertainty*. Nilai *symmetrical uncertainty* berkisar pada rentang 0 sampai dengan 1. *Symmetrical uncertainty* dirumuskan pada persamaan 2.4

$$SU(X, Y) = \frac{IG(X|Y)}{H(X) + H(Y)} \quad (2.4)$$

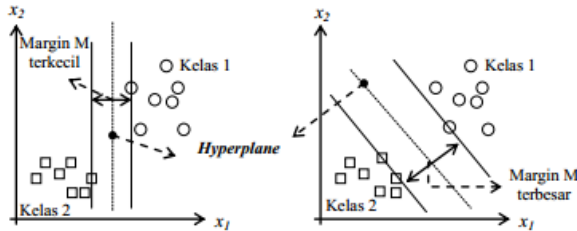
## 2.4 Support Vector Machine

*Support Vector Machine* (SVM) merupakan metode *machine learning* yang melakukan suatu teknik untuk menemukan fungsi pemisah (*classifier*) yang dapat memisahkan data menjadi dua kelas berbeda (Cortes & Vapnik, 1995). Tujuan dari SVM adalah untuk menemukan fungsi pemisah (*hyperplane*) terbaik diantara fungsi yang tidak terbatas jumlahnya untuk memisahkan dua buah kelas. Kelebihan dari metode SVM adalah kemampuan generalisasi, yaitu kemampuan untuk mengklasifikasikan data lain yang tidak termasuk dalam data yang dipakai pada *machine learning* (Sain & Purnami, 2015). SVM untuk klasifikasi dapat bekerja pada kasus klasifikasi linier maupun nonlinier.



### 2.4.1 Klasifikasi SVM Linier

Pada dasarnya, konsep SVM adalah linier *classifier* yaitu klasifikasi ke dalam 2 kelas yang secara linier dapat dipisahkan. Data dapat dipisahkan secara linier berarti terdapat banyak *hyperplane* berbeda yang dapat memisahkan data ke dalam kelas yang berbeda.



Gambar 2.2 Konsep *Hyperplane* pada SVM Linier

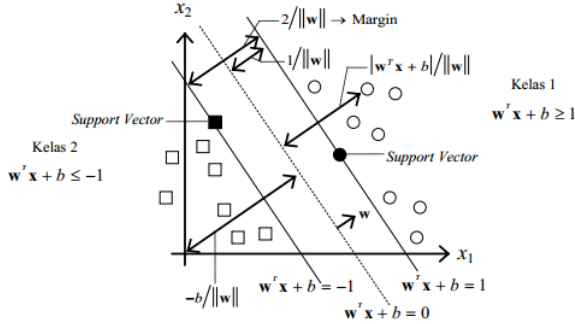
Gambar 2.2 menunjukkan beberapa pengamatan yang merupakan anggota kelas 1 dan 2. *Hyperplane* ditunjukkan oleh garis putus-putus pada gambar, sedangkan margin adalah jarak antara *hyperplane* dengan data yang paling dekat dengan *hyperplane* pada tiap kelas. Kemampuan generalisasi tergantung pada lokasi *hyperplane*, dan *hyperplane* dengan margin terbesar disebut dengan *hyperplane* yang optimal (Cortes & Vapnik, 1995).

Misalkan diberikan suatu himpunan  $\mathbf{x}_i = \{x_1, x_2, \dots, x_n\}$  dimana  $x_i \in R^n, i = 1, 2, \dots, n$ . Diketahui bahwa  $\mathbf{x}_i$  berpola tertentu yaitu ketika  $\mathbf{x}_i$  termasuk dalam suatu *class* (kelas), maka  $\mathbf{x}_i$  diberikan label (target)  $y_i = +1$  dan jika tidak maka diberi label  $y_i = -1$ . Apabila data terpisah secara linier, maka fungsi pemisah atau *hyperplane* didefinisikan oleh (Cortes & Vapnik, 1995):

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2.5)$$

dimana  $\mathbf{w}$  adalah vektor berdimensi  $m$  dan  $b$  adalah suatu konstanta yang disebut bias. Data  $\mathbf{x}_i$  termasuk dalam  $y_i = +1$  bila  $\mathbf{w}^T \mathbf{x}_i + b > 0$  dan data  $\mathbf{x}_i$  termasuk dalam  $y_i = -1$  apabila  $\mathbf{w}^T \mathbf{x}_i + b < 0$ .

Ilustrasi *hyperplane* pemisah dan margin SVM pada data yang dapat dipisahkan secara linier terdapat pada gambar 2.3.



Gambar 2.3 *Hyperplane* SVM Linier

Nilai margin merupakan nilai jarak terdekat *hyperplane* dengan data yang paling dekat dengan *hyperplane* tiap kelas. Jarak antara data  $\mathbf{x}$  pada tiap kelas dengan *hyperplane* adalah pada persamaan (2.6).

$$d(\mathbf{w}, b; \mathbf{x}) = \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|} \quad (2.6)$$

Jarak terdekat antara data  $\mathbf{x}$  dengan *hyperplane* pada kelas 1 dan kelas 2 masing-masing adalah  $\frac{1}{\|\mathbf{w}\|}$ , sehingga nilai margin antara bidang pembatas yaitu  $\frac{2}{\|\mathbf{w}\|}$ , dengan  $\|\mathbf{w}\|$  adalah jarak *Euclidean* (norm *euclidean*) dari  $\mathbf{w}$ . Panjang vektor  $\mathbf{w}$  adalah  $\text{norm} \|\mathbf{w}\| = \sqrt{\mathbf{w}^T \mathbf{w}} = \sqrt{w_1^2 + w_2^2 + \dots + w_p^2}$ .

*Hyperplane* optimum diperoleh dengan memaksimalkan nilai margin. Nilai margin akan maksimal bila nilai  $\|\mathbf{w}\|$  minimal. Secara matematis, formulasi permasalahan optimasi SVM untuk klasifikasi linier dalam primal space adalah:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.7)$$

dengan fungsi kendala  $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, n$ .

Solusi dari permasalahan persamaan kuadrat dengan fungsi batasan berupa pertidaksamaan tersebut dapat diperoleh dengan menggunakan fungsi *Lagrange Multiplier* pada persamaan (2.8),

$$L_p(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^M \alpha_i \left[ y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \right] \quad (2.8)$$

dimana  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)^T$  dan  $\alpha_i$  adalah pengganda fungsi *lagrange* yang bernilai nol atau positif ( $\alpha_i \geq 0$ ). Nilai optimal pada persamaan 2.8 dapat dihitung dengan meminimalkan  $L_p$  terhadap  $\mathbf{w}$  dan  $b$ , serta memaksimalkan  $L_p$  terhadap  $\alpha_i$ .

*Karush-Kuhn-Tucker* (KKT) (Gunn, 1998) kondisi optimal order pertama adalah:

$$\frac{\partial L_p(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0, \quad \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i = 0 \quad (2.9)$$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L_p(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0, \quad 0 - \sum_{i=1}^M \alpha_i y_i \quad (2.10)$$

$$\sum_{i=1}^M \alpha_i y_i = 0$$

substitusi hasil KKT persamaan (2.9) dan (2.10) pada persamaan (2.8), sehingga didapatkan lagrangian untuk dual problem:

$$L_D(\boldsymbol{\alpha}) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j. \quad (2.11)$$

Permasalahan secara dual yaitu memaksimumkan persamaan

(2.11) terhadap  $\alpha_i$  dengan fungsi batasan  $\sum_{i=1}^M y_i \alpha_i = 0$ ,  $\alpha_i \geq 0$

untuk  $i=1, 2, \dots, M$ , sehingga akan menentukan nilai pengganda *lagrange*  $\alpha_i$ . Data yang berasosiasi positif dengan  $\alpha_i$  disebut

*support vector* untuk kelas 1 dan 2. Fungsi *Hyperplane* pemisah yang optimal ditunjukkan pada persamaan (2.12)

$$D(\mathbf{x}) = \sum \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \quad (2.12)$$

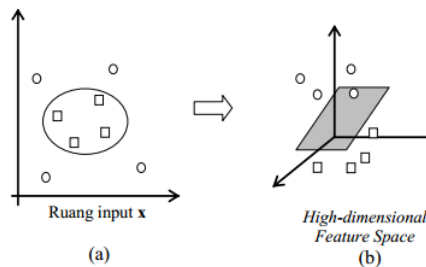
dengan nilai  $b = \frac{1}{S} \sum (y_i - \mathbf{w}^T \mathbf{x}_i)$ .

Selanjutnya data testing  $\mathbf{x}$  akan diklasifikasikan menjadi

$$\mathbf{x} \in \begin{cases} \text{Kelas 1, jika } D(x) > 0 \\ \text{Kelas 2, jika } D(x) < 0 \end{cases} \quad (2.13)$$

#### 2.4.2 Klasifikasi SVM Nonlinier

Pada umumnya masalah dalam dunia nyata jarang yang bersifat linier, akan tetapi kebanyakan bersifat nonlinier. *Hyperplane* yang optimal dalam SVM akan memaksimalkan kemampuan generalisasi. Akan tetapi, apabila data *training* tidak dapat dipisahkan secara linier, maka *classifier* yang diperoleh belum memiliki kemampuan generalisasi yang maksimal meskipun *hyperplane* yang diperoleh sudah optimal. Penyelesaian dalam masalah nonlinier tersebut, SVM dimodifikasi dengan memasukkan fungsi kernel. Ruang input dengan dua dimensi tidak dapat memisahkan data ke dalam dua kelas secara linier, sehingga dilakukan pemetaan vektor input oleh fungsi  $\Phi(\mathbf{x})$  ke ruang vektor baru yang berdimensi lebih tinggi (3 dimensi).



**Gambar 2.2** *Hyperplane Non-Linear*

Pada ilustrasi gambar 2.3 bentuk pemetaan nonlinier tidak perlu diketahui secara eksplisit. Pada SVM nonlinier hal tersebut ditransformasi dalam bentuk  $\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ . Fungsi transformasi

pada SVM menggunakan “Kernel Trick” (Scholkopf & Simola, 2002). Kernel trick adalah menghitung *scalar product* dalam bentuk sebuah fungsi kernel yang dirumuskan pada persamaan (2.14).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (2.14)$$

Fungsi kernel memberikan berbagai kemudahan, karena dengan menggunakan kernel untuk menentukan *support vector* pada SVM tidak perlu untuk mengetahui bentuk dari pemetaan  $\phi(\mathbf{x})$  yang sebenarnya. Fungsi kernel yang umum digunakan pada metode SVM (Scholkopf & Simola, 2002) adalah:

1. Kernel Linier

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

2. Kernel Polynomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (-\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^p, \gamma > 0$$

3. Kernel Radial Basis Function (RBF)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

4. Sigmoid Kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(-\gamma \mathbf{x}_i^T \mathbf{x}_j + r\right)$$

Persamaan dual pada klasifikasi nonlinier diperoleh dari penerapan *kernel trick* pada persamaan (2.15).

$$L_D(\mathbf{a}) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (2.15)$$

Fungsi keputusan pada SVM nonlinier diperoleh melalui persamaan (2.16).

$$D(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b \quad (2.16)$$

dengan  $K(\mathbf{x}_i, \mathbf{x})$  adalah fungsi kernel. Selanjutnya data *testing* diklasifikasikan menggunakan fungsi keputusan pada persamaan (2.17).



$$\mathbf{x} \in \begin{cases} \text{Kelas 1, jika } D(x) > 0 \\ \text{Kelas 2, jika } D(x) < 0 \end{cases} \quad (2.17)$$

Pemilihan fungsi kernel yang tepat merupakan hal yang sangat penting karena akan menentukan *feature* space dimana fungsi *classifier* akan dicari. Menurut Hsu, Chang dan Lin (2004), fungsi kernel yang direkomendasikan untuk diuji pertama kali adalah fungsi kernel RBF karena dapat memetakan hubungan tidak *linear*, RBF lebih robust terhadap outlier karena fungsi kernel RBF berada antara selang  $(-\infty, \infty)$  sedangkan fungsi kernel yang lain memiliki rentang antara  $(-1$  sampai dengan  $1)$ . Selain itu menurut Scholkopf dan Simola (2002) fungsi kernel Gaussian RBF mampu secara otomatis menentukan nilai, lokasi dari *center* dan nilai pembobot dan bisa mencakup rentang tak terhingga. Gaussian RBF juga efektif menghindari *overfitting* dengan memilih nilai yang tepat untuk parameter C dan  $\gamma$  dan RBF baik digunakan ketika tidak ada pengetahuan terdahulu.

## 2.5 Naïve Bayes Classifier

*Naïve Bayes* termasuk ke dalam pembelajaran *supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar yang akan dijadikan label kelas data masukan tersebut. *Naïve Bayes* merupakan perhitungan *teorema bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi dengan asumsi independensi (ketidaktergantungan) yang kuat (naif) antar satu kelas dengan kelas yang lain. Selain itu, algoritma *Naïve Bayes* juga memiliki tingkat akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada *database* yang besar (Han & Kamber, 2006). Secara umum, persamaan *Naïve Bayes* dapat dinotasikan pada persamaan (2.3) (Taheri & Mammadov, 2013):

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (2.18)$$

$P(A|B)$  merupakan *posterior probability* karena nilainya yang bergantung pada nilai  $B$ , sedangkan  $P(A)$  merupakan *prior probability* karena nilainya tidak bergantung pada nilai  $B$ .

Klasifikasi dengan *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri kelas lainnya (Zhang, 2004). Klasifikasi *Naïve Bayes* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas  $y$ . Misalkan diberikan  $x_1, x_2, \dots, x_p$  adalah atribut yang digunakan untuk menentukan kelas  $y$ . Perhitungan *posterior probability* untuk setiap kelas  $y_l$  menggunakan teorema Bayes pada persamaan (2.19) (Taheri & Mammadov, 2013).

$$P(y_l | x_1, \dots, x_p) = \frac{P(y_l)P(x_1, \dots, x_p | y_l)}{P(x_1, \dots, x_p)} \quad (2.19)$$

Persamaan 2.16 diperlukan perhitungan  $P(x_1, \dots, x_p | y_l)$  dan setiap atribut diasumsikan saling bebas untuk kelas  $y$ , maka:

$$P(x_1, x_2, \dots, x_p) = P(x_1 | y)P(x_2 | y) \dots (x_p | y) \quad (2.20)$$

Jika terdapat atribut yang bersifat kuantitatif atau kontinyu, maka  $P(x_i | y_l)$  dihitung menggunakan pendekatan distribusi normal.

$$P(x_i | y_l) = \frac{1}{\sigma_{il}\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu_{il})^2}{2\sigma_{il}^2}\right) \quad (2.21)$$

Estimasi peluang  $P(x_i | y_l)$  dapat dihitung pada setiap atribut  $x_i$  dan  $y_l$ , sehingga data baru akan diklasifikasikan ke dalam kelas  $y_l$  apabila peluang yang didapat lebih besar dari yang lain.

## 2.6 Pengukuran Akurasi Klasifikasi

Nilai akurasi hasil klasifikasi dapat dihitung menggunakan *confusion matrix*. *Confusion matrix* adalah alat yang digunakan untuk menganalisis seberapa baik *classifier* mengenali data dari kelas yang berbeda.

**Tabel 2.1** *confusion matrix*

Aktual	Prediksi	
	$\hat{\pi}_1$	$\hat{\pi}_2$
$\hat{\pi}_1$	$n_{11}$	$n_{12}$
$\hat{\pi}_2$	$n_{21}$	$n_{22}$

keterangan:

$n_{11}$  : jumlah prediksi  $\hat{\pi}_1$  yang tepat diklasifikasikan ke  $\hat{\pi}_1$

$n_{22}$  : jumlah prediksi  $\hat{\pi}_2$  yang tepat diklasifikasikan ke  $\hat{\pi}_2$

$n_{21}$  : jumlah prediksi  $\hat{\pi}_1$  yang tepat diklasifikasikan ke  $\hat{\pi}_2$

$n_{12}$  : jumlah prediksi  $\hat{\pi}_2$  yang tepat diklasifikasikan ke  $\hat{\pi}_1$

*Confusion matrix* pada tabel 2.1 digunakan untuk beberapa pengukuran performa klasifikasi yaitu akurasi, sensitivitas, spesifitas, dan presisi (Han & Kamber, 2006).

$$akurasi = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{22} + n_{21}} \quad (2.22)$$

$$sensitifitas = \frac{n_{11}}{n_{11} + n_{12}} \quad (2.23)$$

$$spesifitas = \frac{n_{22}}{n_{22} + n_{21}} \quad (2.24)$$

Akurasi menunjukkan efektifitas *classifier* yaitu secara menyeluruh. Semakin besar nilai akurasi maka semakin baik kinerja *classifier*. Sensitivitas digunakan untuk mengukur sebuah *classifier* untuk mengidentifikasi kelas positif, sedangkan untuk spesifikasi mengukur efektifitas *classifier* dalam mengidentifikasi kelas negatif.

Perhitungan akurasi klasifikasi dengan *confusion matrix* dapat digunakan untuk komposisi data yang *balance*. Akurasi klasifikasi data yang *unbalance* dapat dihitung menggunakan

*Geometric-mean (G-mean)* dan *Area Under ROC Curve (AUC)* (Bekkar, Djemaa, & Alitouch, 2013).

$$Gmean = \left( \prod_{i=1}^g R_i \right) = \sqrt{\text{sensitifitas} \times \text{spesifitas}} \quad (2.25)$$

$$AUC = \frac{1}{g} \sum_{i=1}^g R_i = \frac{1}{2} (\text{sensitifitas} + \text{spesifitas}) \quad (2.26)$$

dengan  $R_i = \frac{n_{ii}}{\sum_{l=1}^g n_{il}}$  dan  $i = 1, 2, \dots, g$

keterangan:

$n_{ii}$  : jumlah prediksi  $\hat{\pi}_i$  yang tepat diklasifikasikan ke  $\hat{\pi}_i$

$n_{il}$  : jumlah prediksi  $\hat{\pi}_l$  yang tepat diklasifikasikan ke  $\hat{\pi}_i$

## 2.7 Energi Listrik Sektor Rumah Tangga

PT Perusahaan Listrik Negara (Persero) adalah perusahaan yang berusaha di bidang tenaga listrik untuk kepentingan masyarakat dan negara (Siregar & Warman, 2013). Pelanggan PT PLN (Persero) dibagi menjadi 5 golongan antara lain sektor rumah tangga, industri, bisnis, sosial, dan penerangan umum.

**Tabel 2.2** Jumlah Pelanggan Rumah Tangga PLN Kota Surabaya

Golongan Daya	Keterangan	Jumlah pelanggan		
		SBB	SBU	SBS
R-1/450 VA	Subsidi	42.531	25.579	44.667
R-1/900 VA	Subsidi	11.984	39.946	28.218
R-1/900 VA-RTM	Non subsidi	83.830	110.973	161.520
R-1/1.300 VA	Non subsidi	59.957	70.874	137.908
R-1/2.200 VA	Non subsidi	15.054	31.734	67.913
R-2/3.500 VA s.d 5.500 VA	Non subsidi	5.667	16.664	40.978
R-3/6.600 VA keatas	Non subsidi	439	5.251	12.903

Sumber: PT.PLN (Persero) Distribusi Jawa Timur, Desember 2017.

Pelanggan rumah tangga dengan daya 900 VA dibagi menjadi golongan R-1/900 VA dan golongan R-1/900 VA-RTM (pelanggan rumah tangga mampu non subsidi).

## BAB III METODE PENELITIAN

### 3.1 Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari PT. PLN (Persero) Distribusi Jawa Timur yaitu data pelanggan rumah tangga bulan Januari 2018 yang memperoleh subsidi listrik, sedangkan data sosial ekonomi rumah tangga tersebut diperoleh dari Basis Data Terpadu (BDT) program perlindungan sosial yang dikelola oleh Badan Perencanaan Kota Surabaya. Data yang diperoleh dari PLN hanya berupa data pelanggan subsidi 900 VA, selanjutnya data ini dilakukan penyesuaian dengan data fakir miskin dari Bappeko dan diperoleh data pelanggan yang sesuai sebanyak 28949 pelanggan. Sedangkan data pelanggan rumah tangga subsidi 450 VA seluruhnya diambil langsung dari data fakir miskin Bappeko dan diperoleh sebanyak 50315 pelanggan. Total pelanggan rumah tangga yang akan digunakan pada penelitian ini berjumlah 79264 pelanggan. Selanjutnya, data dibagi menjadi data *testing* dan data *training* yang dilakukan secara acak menggunakan metode *k-fold cross validation* dengan melakukan proses *folding* sebanyak 10 kali. Pada setiap pengujian, 1 *subset* digunakan sebagai data *testing* dan *subset* yang lain digunakan sebagai data *training*.

**Tabel 3.3** Struktur Data Penelitian

Sampel	Kelas	$Y$	$X_1$	$X_2$	...	$X_{16}$
1	-1	$Y_{1,-1}$	$X_{1,-1,1}$	$X_{1,-1,2}$	...	$X_{1,-1,16}$
2	-1	$Y_{2,-1}$	$X_{2,-1,1}$	$X_{2,-1,2}$	...	$X_{2,-1,16}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
$n_1$	-1	$Y_{n_1,-1}$	$X_{n_1,-1,1}$	$X_{n_1,-1,2}$	...	$X_{n_1,-1,16}$
1	1	$Y_{1,1}$	$X_{1,1,1}$	$X_{1,1,2}$	...	$X_{1,1,16}$
2	1	$Y_{2,1}$	$X_{2,1,1}$	$X_{2,1,2}$	...	$X_{2,1,16}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
$n_2$	1	$Y_{n_2,1}$	$X_{n_2,1,1}$	$X_{n_2,1,2}$	...	$X_{n_2,1,16}$



Keterangan:

$n_1$  = jumlah rumah tangga pada kategori kelas subsidi 450 VA  
yaitu sebanyak 50315 pelanggan.

$n_2$  = jumlah rumah tangga pada kategori kelas subsidi 900 VA  
yaitu sebanyak 28949 pelanggan.

### 3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini disajikan dalam tabel 3.1. Variabel respon yang digunakan terdiri dari 2 kategori yaitu pelanggan rumah tangga subsidi listrik 450 VA dan pelanggan subsidi 900 VA.

**Tabel 3.2** Variabel Penelitian

Variabel	Nama variabel	Label	Skala
Y	Daya listrik	1. subsidi 450 VA 2. subsidi 900 VA	Ordinal
$X_1$	Status kepemilikan tempat tinggal	1. bebas sewa/dinas 2. kontrak/sewa 3. milik sendiri	Ordinal
$X_2$	Luas bangunan	1. $\leq 36 \text{ m}^2$ 2. $36 \text{ m}^2 < \text{luas} \leq 70 \text{ m}^2$ 3. $> 70 \text{ m}^2$	Ordinal
$X_3$	Jenis lantai terluas	1. tanah 2. semen/bata merah 3. ubin/tegel/teraso 4. keramik 5. marmer/granit	Ordinal
$X_4$	Jenis dinding terluas	1. tembok 2. kayu 3. bambu	Ordinal
$X_5$	Jenis atap terluas	1. seng 2. asbes 3. genteng tanah liat 4. genteng metal 5. genteng keramik 6. beton/genteng beton	Ordinal
$X_6$	Jumlah kamar tidur	1. jumlah 0 – 1 kamar 2. jumlah 2 – 3 kamar 3. jumlah 4 – 5 kamar	Ordinal

**Tabel 3.3** Variabel Penelitian (Lanjutan)

$X_7$	Sumber air minum	1. Sumur terlindung 2. Sumur bor/pompa 3. Leding eceran 4. Leding meteran 5. Air isi ulang 6. Air kemasan bermerk	Ordinal
$X_8$	Bahan bakar memasak	1. Listrik 2. Gas >3kg 3. Gas 3kg 4. Gas kota/biogas 5. Minyak tanah 6. Kayu bakar	Ordinal
$X_9$	Kepemilikan lemari es	1. Tidak 2. Ya	Ordinal
$X_{10}$	Kepemilikan televisi	1. Tidak 2. Ya	Ordinal
$X_{11}$	Kepemilikan komputer/laptop	1. Tidak 2. Ya	Ordinal
$X_{12}$	Kepemilikan sepeda	1. Tidak 2. Ya	Ordinal
$X_{13}$	Kepemilikan sepeda motor	1. Tidak 2. Ya	Ordinal
$X_{14}$	Peserta program raskin	1. Ya 2. Tidak	Ordinal
$X_{15}$	Jumlah anggota keluarga	1. anggota 1 – 2 orang 2. anggota 3 – 4 orang 3. anggota 5 – 6 orang 4. anggota $\geq 7$ orang	Ordinal
$X_{16}$	Status kesejahteraan	1. kondisi kesejahteraan 10% terendah 2. kondisi kesejahteraan 11% hingga 20% terendah 3. kondisi kesejahteraan 21% hingga 30% terendah 4. kondisi kesejahteraan 31% hingga 40% terendah	Ordinal

### 3.3 Langkah Analisis

Langkah-langkah analisis yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Penyesuaian data PT PLN dan Bappeko Surabaya  
Data yang diperoleh dari PT PLN yaitu data pelanggan rumah tangga bulan Januari 2018 yang memperoleh subsidi listrik, dan dilakukan penyesuaian dengan data fakir miskin dari Basis Data Terpadu (BDT) program perlindungan sosial yang dikelola oleh Badan Perencanaan Kota Surabaya untuk mendapatkan data sosial ekonomi rumah tangga tersebut. Setelah dilakukan penyesuaian data, diperoleh data pelanggan rumah tangga subsidi listrik 900 VA yang sesuai antara data PLN dan Bappeko Surabaya sebanyak 50315 pelanggan, sedangkan untuk seluruh data pelanggan rumah tangga subsidi 450VA diperoleh dari data Bappeko Surabaya tanpa proses penyesuaian dengan data PLN, yaitu berjumlah 28949 pelanggan, sehingga total pelanggan rumah tangga yang akan digunakan pada penelitian ini berjumlah 79264 pelanggan. Data pelanggan rumah tangga tersebut kemudian diubah menurut keperluan algoritma dan jenis tipe data.
2. Menganalisis karakteristik data pelanggan rumah tangga yang memperoleh subsidi listrik daya 450 VA dan pelanggan subsidi 900 VA.
3. Melakukan seleksi variabel menggunakan metode FCBF, sehingga terdapat 2 kelompok data yaitu data menggunakan seluruh variabel prediktor dan data hanya menggunakan variabel yang terseleksi. Selanjutnya 2 kelompok data tersebut masing-masing dibagi menjadi data *training* dan data *testing*.
4. Pembagian data  
Pembagian data *testing* dan data *training* menggunakan *k-fold cross validation* yaitu sebanyak 10 *fold*. Pada setiap pengujian, 1 *subset* digunakan sebagai data *testing* sebanyak 7927 data, sedangkan *subset* yang lain digunakan sebagai data *training* sebanyak  $79264 - 7927 = 71337$  data.

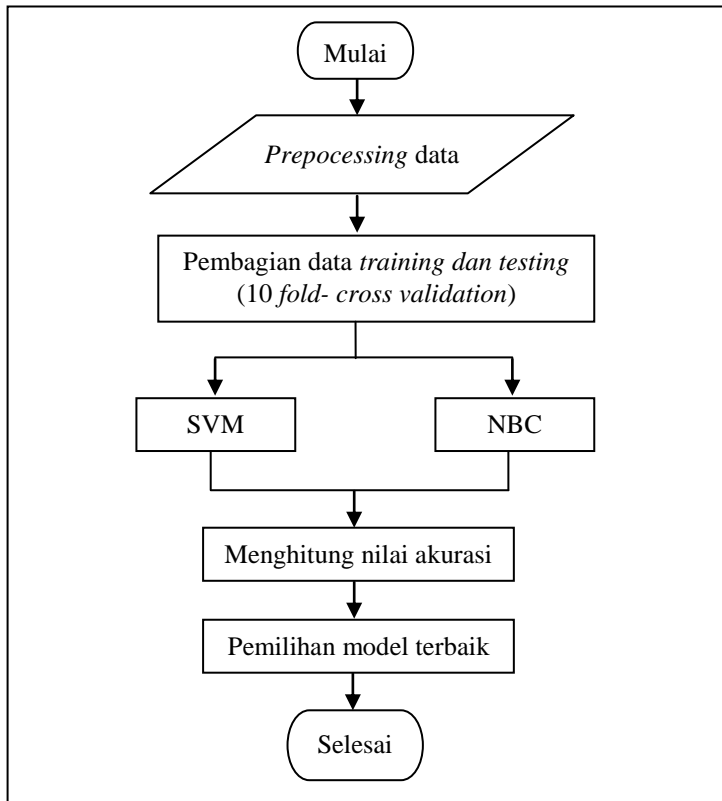
5. Melakukan klasifikasi dengan metode *Support Vector Machine* melalui 2 tahapan yaitu menggunakan semua variabel prediktor dan menggunakan seleksi variabel dengan metode FCBF. Langkah-langkah menggunakan metode *Support Vector Machine* adalah sebagai berikut:
  - a. Menentukan fungsi kernel yang akan digunakan yaitu pada penelitian ini menggunakan fungsi kernel Linier dan *Radial Basis Function* (RBF). Pada kernel Linier terdapat parameter cost ( $C$ ) dan pada kernel RBF terdapat dua parameter yang diperlukan yaitu parameter cost ( $C$ ) dan gamma ( $\gamma$ ).
  - b. Menentukan *range* nilai parameter  $C$  dan  $\gamma$  yaitu pada penelitian ini menggunakan nilai  $C = 0,1, 1, 10$ , dan  $100$  dan nilai  $\gamma = 0,1, 1, 10$ , dan  $100$  berdasarkan penelitian yang dilakukan oleh Sahitayakti (2015).
  - c. Melakukan klasifikasi SVM pada 10 *fold* data *training*. Pada masing-masing *fold* data *training* untuk kernel linier dilakukan 5 kombinasi nilai  $C$ , sehingga total terdapat 50 kombinasi, sedangkan untuk kernel RBF pada masing-masing *fold* data *training* dilakukan sebanyak 16 kombinasi nilai  $C$  dan  $\gamma$ , sehingga total terdapat 160 kombinasi.
  - d. Menghitung nilai akurasi, AUC, dan *G-Mean* pada masing-masing *fold* hingga semua kombinasi selesai.
  - e. Menghitung rata-rata nilai akurasi, AUC, dan *G-Mean* yang diperoleh dari klasifikasi SVM pada 10 *fold* data *training*.
  - f. Menentukan nilai  $C$  dan  $\gamma$  yang paling optimal yaitu rata-rata nilai akurasi, AUC, dan *G-Mean* yang paling tinggi (Huang & Wang, 2006).
  - g. Nilai  $C$  dan  $\gamma$  yang paling optimal kemudian diuji pada 10 *fold* data testing dan performa klasifikasi ditentukan berdasarkan nilai rata-rata akurasi, AUC, dan *G-Mean* pada 10 *fold* data testing tersebut (Huang & Wang, 2006).

6. Melakukan klasifikasi menggunakan *Naïve Bayes Classifier* melalui 2 tahapan yaitu menggunakan semua variabel prediktor dan menggunakan seleksi variabel dengan metode FCBF. Langkah-langkah menggunakan metode *Naïve Bayes Classifier* adalah sebagai berikut:
  - a. Menghitung nilai probabilitas masing-masing kategori yaitu probabilitas pelanggan rumah tangga subsidi 450 VA dan 900 VA.
  - b. Menghitung probabilitas semua kategori yang ada pada variabel prediktor terhadap kategori pelanggan rumah tangga subsidi 450 VA dan 900 VA.
  - c. Model probabilitas NBC disimpan dan digunakan untuk tahap data *testing*.
  - d. Menghitung probabilitas semua kategori yang diujikan
  - e. Menentukan kelompok kategori berdasarkan nilai probabilitas yang tertinggi.
7. Membandingkan performansi antara metode *Support Vector Machine* dan *Naïve Bayes Classifier* berdasarkan nilai akurasi, AUC, dan *G-Mean* yang tertinggi.



### 3.4 Diagram Alir

Diagram alir berdasarkan langkah analisis yang digunakan dalam penelitian ini ditunjukkan pada Gambar 3.1.



**Gambar 3.1** Diagram Alir Penelitian

*(Halaman ini sengaja dikosongkan)*

## BAB IV ANALISIS DAN PEMBAHASAN

Pada bab analisis dan pembahasan, terdapat empat subbab yang akan dibahas. Subbab pertama akan dijelaskan mengenai karakteristik rumah tangga yang memperoleh subsidi listrik di Kota Surabaya. Subbab kedua merupakan pembahasan klasifikasi rumah tangga yang memperoleh subsidi listrik dengan metode *Support Vector Machine*. Pada subbab ketiga dijelaskan mengenai klasifikasi rumah tangga yang memperoleh subsidi listrik dengan metode *Naïve Bayes Classifier*. Pada subbab terakhir, subbab keempat akan dibahas perbandingan dari hasil klasifikasi rumah tangga yang memperoleh subsidi listrik di Kota Surabaya dengan metode *Support Vector Machine* dan *Naïve Bayes Classifier*.

### 4.1 Karakteristik Rumah Tangga Subsidi Listrik

Penyediaan dana bagi kelompok masyarakat miskin dan tidak mampu diberikan dalam bentuk subsidi terhadap tarif tenaga listrik untuk pelanggan rumah tangga dengan daya listrik 450 VA dan 900 VA. Pada penelitian ini, terdapat 16 variabel yang mencerminkan karakteristik rumah tangga yang memperoleh subsidi listrik berdasarkan Basis Data Terpadu untuk yang dikelola oleh Badan Perencanaan Kota Surabaya.

**Tabel 4.1** Karakteristik Berdasarkan Daya Listrik  
dan Jumlah Anggota Keluarga

Anggota keluarga	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
1 – 2 orang	15380	30,6	8745	30,2
<b>3 – 4 orang</b>	<b>24045</b>	<b>47,8</b>	<b>13590</b>	<b>46,9</b>
5 – 6 orang	9411	18,7	5705	19,7
> 7 Orang	1479	2,9	909	3,1
Total	50315	100%	28949	100%

Tabel 4.1 menyajikan hasil *cross tabulation* variabel jumlah anggota keluarga dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.1 diketahui bahwa sebagian besar pelanggan rumah tangga memiliki jumlah anggota keluarga sebanyak 3 atau 4 orang untuk kategori pelanggan subsidi 450 VA dan 900 VA.

**Tabel 4.2** Karakteristik Berdasarkan Daya Listrik dan Status Kepemilikan Bangunan Tempat Tinggal

Kepemilikan	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Dinas/bebas kontrak	8769	17,4	2863	9,9
Kontrak/sewa	13772	27,4	4495	15,5
<b>Milik sendiri</b>	<b>27774</b>	<b>55,2</b>	<b>21591</b>	<b>74,6</b>

Tabel 4.2 menyajikan hasil *cross tabulation* variabel status kepemilikan bangunan tempat tinggal dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.2, sebagian besar pelanggan subsidi 900 VA sebanyak 74,6% pelanggannya sudah memiliki status kepemilikan bangunan milik sendiri. Pelanggan subsidi daya 450 VA masih mencapai 55,2% tempat tinggal dengan status milik sendiri dan sisanya masih berstatus kepemilikan kontrak atau sewa, bebas kontrak, serta tempat tinggal dinas.

**Tabel 4.3** Karakteristik Berdasarkan Daya Listrik dan Luas Lantai

Luas lantai	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
$\leq 36 \text{ m}^2$	<b>33147</b>	<b>65,9</b>	<b>13671</b>	<b>47,2</b>
$>36 \text{ m}^2$ dan $\leq 70 \text{ m}^2$	12030	23,9	10275	35,5
$>70 \text{ m}^2$	5138	10,2	5003	17,3

Tabel 4.3 menyajikan hasil *cross tabulation* variabel lantai terluas dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.3, sebagian besar pelanggan 450 VA mencapai 65,9% dari pelanggannya memiliki luas lantai yang  $\leq 36 \text{ m}^2$ , sedangkan untuk pelanggan subsidi daya 900 VA masih banyak pelanggan yang memiliki luas lantai  $\leq 36 \text{ m}^2$  yaitu sekitar 47,2% pelanggan, namun 35,5% pelanggannya juga sudah banyak memiliki luas lantai berkisar antara  $36 \text{ m}^2$  dan  $70 \text{ m}^2$  dan 17,3% sisanya memiliki luas lantai lebih dari  $70 \text{ m}^2$ .

Tabel 4.4 menyajikan hasil *cross tabulation* variabel jenis lantai terluas dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.4, sebagian besar pelanggan subsidi 900 VA sebanyak 61,7% pelanggannya sudah memiliki jenis lantai keramik, sedangkan pelanggan subsidi daya 450 VA sebagian besar sudah menggunakan keramik, namun masih mencapai 44,8% dari total

pelanggannya. Urutan kedua dan ketiga pelanggan subsidi daya 450 VA juga masih banyak menggunakan jenis lantai yang terbuat dari ubin/tegel serta dari semen/bata merah.

**Tabel 4.4** Karakteristik Berdasarkan Daya Listrik dan Jenis Lantai Terluas

Jenis lantai	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Tanah	751	1,5	230	0,8
Semen/bata merah	11150	22,2	4144	14,3
Ubin/tegel/teraso	15678	31,2	6562	22,7
<b>Keramik</b>	<b>22550</b>	<b>44,8</b>	<b>17851</b>	<b>61,7</b>
Marmer/granit	186	0,4	162	0,6

Tabel 4.5 menyajikan hasil *cross tabulation* variabel jenis dinding terluas dengan kategori daya listrik. Berdasarkan Tabel 4.5, sebagian besar pelanggan daya 450 VA dan 900 VA subsidi sudah menggunakan jenis dinding tembok yaitu masing-masing dengan persentase sebesar 86,2% dan 91,7 %.

**Tabel 4.5** Karakteristik Berdasarkan Daya Listrik dan Jenis Dinding Terluas

Jenis dinding	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Bambu	1921	3,8	1005	3,5
Kayu	5013	10,0	1387	4,8
<b>Tembok</b>	<b>43381</b>	<b>86,2</b>	<b>26557</b>	<b>91,7</b>

**Tabel 4.6** Karakteristik Berdasarkan Daya Listrik dan Jenis Atap Terluas

Jenis atap	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Seng	1950	3,9	752	2,6
Asbes	13128	26,1	7802	27,0
<b>Tanah liat</b>	<b>33891</b>	<b>67,4</b>	<b>19398</b>	<b>67,0</b>
Metal	233	0,5	336	1,2
Keramin	249	0,5	107	0,4
Beton	864	1,7	554	1,9

Tabel 4.6 menyajikan hasil *cross tabulation* variabel jenis atap terluas dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.6, sebagian besar pelanggan daya 450 VA dan 900 VA subsidi menggunakan jenis atap yaitu genteng tanah liat dengan persentase masing-masing daya sudah mencapai 67%. Jenis atap

terbanyak kedua setelah genteng tanah liat yang juga banyak digunakan pelanggan subsidi daya 450 VA dan pelanggan subsidi 900 VA adalah jenis asbes.

**Tabel 4.7** Karakteristik Berdasarkan Daya Listrik dan Jumlah Kamar Tidur

Jumlah	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
<b>0 – 1</b>	<b>32262</b>	<b>64,1</b>	11542	39,9
<b>2 – 3</b>	17521	34,8	<b>16636</b>	<b>57,5</b>
> 3	532	1,1	771	2,7

Tabel 4.7 menyajikan hasil *cross tabulation* variabel jumlah kamar tidur dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.7, pelanggan subsidi daya 450 VA dengan persentase 64,1% memiliki jumlah kamar tidur hanya sebanyak 1 kamar atau tidak memiliki kamar tidur dalam tempat tinggalnya, sedangkan untuk kategori pelanggan 900 VA subsidi dengan persentase paling tinggi sebesar 57,5% memiliki jumlah kamar tidur sebanyak 2 atau 3 kamar.

**Tabel 4.8** Karakteristik Berdasarkan Daya Listrik dan Sumber Air Minum

Sumber	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Sumur terlindung	875	1,7	369	1,3
Sumur bor/pompa	304	0,6	128	0,4
Leding eceran	8793	17,5	3891	13,4
Leding meteran	14591	29,0	8286	28,6
<b>Air isi ulang</b>	<b>24148</b>	<b>48,0</b>	<b>13935</b>	<b>48,1</b>
Air kemasan bermerk	1604	3,2	2340	8,1

Tabel 4.8 menyajikan hasil *cross tabulation* variabel sumber air minum dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.8, sebagian besar pelanggan daya 450 VA dan 900 VA menggunakan sumber air minum yaitu air isi ulang dengan persentase masing-masing kategori hampir sama yaitu 48%. Persentase sumber air minum pada urutan selanjutnya yang paling banyak digunakan yaitu pelanggan subsidi 450 VA dan 900 VA juga menggunakan leding meteran dan leding eceran sebagai sumber air minum.



**Tabel 4.9** Karakteristik Berdasarkan Daya Listrik dan Bahan Bakar untuk Memasak

Sumber	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Kayu bakar	299	0,6	272	0,9
Minyak tanah	1468	2,9	408	1,4
Gas kota/biogas	215	0,4	193	0,7
<b>Gas 3 kg</b>	<b>46067</b>	<b>91,6</b>	<b>27184</b>	<b>93,9</b>
Gas >3 kg	969	1,9	315	1,1
Listrik	1297	2,6	577	2,0

Tabel 4.9 menyajikan hasil *cross tabulation* variabel bahan bakar untuk memasak dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.9, sebagian besar pelanggan daya 450 VA dan 900 VA subsidi menggunakan gas 3 kg untuk memasak dengan persentase pelanggan subsidi daya 450 VA dan 900 VA sudah mencapai diatas 90% dari total pelanggannya.

**Tabel 4.10** Karakteristik Berdasarkan Daya Listrik dan Kepemilikan lemari es

Kepemilikan	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
<b>Tidak</b>	<b>34183</b>	<b>67,9</b>	13755	47,5
<b>Ya</b>	16132	32,1	<b>15194</b>	<b>52,5</b>

Tabel 4.10 menyajikan hasil *cross tabulation* variabel kepemilikan lemari es dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.10, pelanggan daya 450 VA sebagian besar tidak memiliki lemari es dengan persentase 67,9%. Sedangkan untuk pelanggan daya 900 VA subsidi sebesar 52,5% dari total pelanggannya sudah memiliki lemari es, namun masih terdapat 47,5% total pelanggannya yang belum memiliki lemari es.

**Tabel 4.11** Karakteristik Berdasarkan Daya Listrik dan Kepemilikan Televisi

Kepemilikan	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Tidak	6474	12,9	2479	8,6
<b>Ya</b>	<b>43841</b>	<b>87,1</b>	<b>26470</b>	<b>91,4</b>

Tabel 4.11 menyajikan hasil *cross tabulation* variabel kepemilikan televisi dengan kelas atau kategori daya listrik.

Berdasarkan Tabel 4.11, sebagian besar pelanggan subsidi daya 450 VA dan 900 VA sudah memiliki televisi dengan persentase hampir mencapai 90% dari total pelanggannya untuk masing-masing kategori.

**Tabel 4.12** Karakteristik Berdasarkan Daya Listrik dan Kepemilikan Komputer/Laptop

Kepemilikan	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
<b>Tidak</b>	<b>49258</b>	<b>97,9</b>	<b>27328</b>	<b>94,4</b>
<b>Ya</b>	1057	2,1	1621	5,6

Tabel 4.12 menyajikan hasil *cross tabulation* variabel kepemilikan komputer/laptop dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.12, sebagian besar pelanggan subsidi daya 450 VA dan 900 VA tidak memiliki komputer/laptop dengan persentase diatas 90% dari total pelanggannya untuk masing-masing kategori.

**Tabel 4.13** Karakteristik Berdasarkan Daya Listrik dan Kepemilikan Sepeda

Kepemilikan	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
<b>Tidak</b>	<b>31408</b>	<b>62,4</b>	12643	43,7
<b>Ya</b>	18907	37,6	<b>16306</b>	<b>56,3</b>

Tabel 4.13 menyajikan hasil *cross tabulation* variabel kepemilikan sepeda dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.13, sebagian besar pelanggan daya 450 VA dengan persentase 62,4% tidak memiliki sepeda, sedangkan pelanggan subsidi daya 900 VA sudah memiliki sepeda dengan persentase pelanggan 56,3%.

**Tabel 4.14** Karakteristik Berdasarkan Daya Listrik dan Kepemilikan Sepeda Motor

Kepemilikan	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Tidak	22402	44,5	10347	35,7
<b>Ya</b>	<b>27913</b>	<b>55,5</b>	<b>18602</b>	<b>64,3</b>

Tabel 4.14 menyajikan hasil *cross tabulation* variabel kepemilikan sepeda motor dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.14, pelanggan subsidi daya 450 VA dengan

persentase 55,5% memiliki sepeda motor, namun masih terdapat 45,5% dari total pelanggannya yang belum memiliki sepeda motor, sedangkan untuk pelanggan subsidi daya 900 VA sudah memiliki sepeda motor dengan persentase 64,3%.

**Tabel 4.15** Karakteristik Berdasarkan Daya Listrik dan Peserta Program Raskin

Kepemilikan	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
Tidak	19857	39,5	12556	43,4
<b>Ya</b>	<b>30458</b>	<b>60,5</b>	<b>16393</b>	<b>56,6</b>

Tabel 4.15 menyajikan hasil *cross tabulation* variabel peserta program raskin dengan kelas atau kategori daya listrik. Berdasarkan Tabel 4.15, sebagian besar pelanggan daya 450 VA dengan persentase 60,5% merupakan peserta program raskin, sedangkan untuk pelanggan daya 900 VA sebesar 56,6% merupakan peserta program raskin dan 43,4% pelanggannya sudah tidak mendapatkan program raskin dari pemerintah.

**Tabel 4.16** Karakteristik Berdasarkan Daya Listrik dan Status Kesejahteraan

Status	450 VA		900 VA subsidi	
	Jumlah	%	Jumlah	%
10% terendah	12249	24,3	3394	11,7
<b>11% – 20% terendah</b>	<b>14169</b>	<b>28,2</b>	7100	24,5
21% – 30% terendah	12918	25,7	8413	29,1
<b>31% – 40 % terendah</b>	10979	21,8	<b>10042</b>	<b>34,7</b>

Tabel 4.16 menyajikan hasil *cross tabulation* variabel kelas atau kategori daya listrik dan status kesejahteraan. Berdasarkan Tabel 4.16, jumlah pelanggan daya 450 VA hampir merata pada masing-masing kategori dengan persentase paling tinggi sebesar 28,2% dengan status kesejahteraan 11% - 20% terendah, sedangkan untuk kategori pelanggan 900 VA subsidi dengan persentase paling tinggi sebesar 34,7% berada pada status kesejahteraan 31% - 40% terendah.

## 4.2 Seleksi Variabel Menggunakan FCBF

Seleksi variabel yang dilakukan menggunakan metode FCBF yaitu dengan melihat nilai dari *Information Gain*. Variabel yang memiliki nilai *Information Gain* paling tinggi merupakan variabel dengan pengaruh yang paling kuat terhadap pembentukan klasifikasi.

**Tabel 4.17** Hasil Seleksi Variabel

No	Variabel	<i>Information Gain</i>
1	Jumlah kamar tidur	0,040
2	Kepemilikan lemari es	0,029
3	Status kepemilikan tempat tinggal	0,027
4	Status kesejahteraan	0,026
5	Luas lantai	0,024
6	Kepemilikan sepeda	0,023
7	Jenis lantai terluas	0,019

Tabel 4.17 menunjukkan bahwa terdapat 7 variabel yang terseleksi menggunakan metode FCBF dari total variabel prediktor sebanyak 16 variabel. Variabel yang terpilih tersebut memberikan pengaruh yang paling kuat dalam pembentukan klasifikasi dibandingkan variabel lain yang tidak terseleksi. Variabel yang paling berpengaruh terhadap pembentukan klasifikasi adalah jumlah kamar tidur. Hasil karakteristik yang disajikan pada Tabel 4.1 hingga 4.16 diperoleh variabel yang membedakan antara karakteristik pelanggan subsidi 450 VA dan 900 VA adalah variabel jumlah kamar tidur, kepemilikan lemari es, kepemilikan sepeda, dan status kesejahteraan, sedangkan hasil seleksi menggunakan FCBF juga diperoleh variabel-variabel tersebut yang berpengaruh terhadap hasil pengelompokan.

## 4.3 Klasifikasi Menggunakan *Support Vector Machine* (SVM)

Metode pertama yang akan digunakan untuk melakukan klasifikasi pelanggan subsidi listrik rumah tangga adalah metode SVM. Pada penelitian ini, metode SVM menggunakan fungsi kernel Linier dan kernel *Radial Basis Function* (RBF). Kernel linier membutuhkan parameter *Cost* (*C*), sedangkan kernel RBF

membutuhkan parameter *Cost* ( $C$ ) dan *Gamma* ( $\gamma$ ). Klasifikasi pada metode SVM ini akan dilakukan 2 tahapan yaitu klasifikasi menggunakan seluruh variabel prediktor, serta klasifikasi dengan variabel hasil seleksi FCBF.

#### 4.2.1 Klasifikasi *Support Vector Machine* Menggunakan Seluruh Variabel Prediktor

Pengaturan parameter  $C$  dan  $\gamma$  untuk membentuk model SVM dilakukan pada data *training*. Parameter  $C$  yang digunakan pada kernel linier yaitu menggunakan nilai  $C = 0,01, 0,1, 1, 10$ , dan  $100$ . Klasifikasi SVM pada kernel linier dilakukan pada  $10$  *fold* data *training* dengan kombinasi nilai parameter  $C$  sebanyak  $5$  kali pada masing-masing *fold*, sehingga total terdapat  $50$  kombinasi. Setiap percobaan pada  $10$  *fold* data *training* menghasilkan nilai akurasi, AUC, dan *G-Mean* yang berbeda-beda, sehingga parameter  $C$  yang optimal adalah parameter yang memberikan nilai rata-rata akurasi, AUC, dan *G-Mean* yang paling tinggi dari percobaan pada  $10$  *fold* data *training*. Pemilihan parameter optimal berdasarkan nilai rata-rata akurasi, AUC, dan *G-Mean* yang paling tinggi pernah dilakukan oleh Huang & Wang (2006).

**Tabel 4.18** Rata-rata Akurasi, AUC, dan *G-Mean* Kernel Linier (*Training*)

Cost ( $C$ )	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
0,01	69,0	63,2	59,5
<b>0,1</b>	<b>69,0</b>	<b>63,3</b>	<b>59,6</b>
<b>1</b>	<b>69,0</b>	<b>63,3</b>	<b>59,6</b>
10	69,0	63,2	59,5
100	68,9	63,1	59,5

Hasil dari Tabel 4.18 dapat diketahui bahwa nilai rata-rata akurasi, AUC, dan *G-Mean* paling tinggi terdapat pada nilai  $C = 0,1$  dan  $C = 1$ . Selanjutnya dipilih nilai  $C = 0,1$  sebagai parameter optimal dan diterapkan pada  $10$  data *testing* pada Tabel 4.19. Dengan menerapkan model optimal pada data *testing*, maka akan diperoleh performa klasifikasi berdasarkan rata-rata dari nilai akurasi, AUC, dan *G-Mean* (Huang & Wang, 2006).

**Tabel 4.19** Nilai Akurasi, AUC, dan *G-Mean* Kernel Linier (*Testing*)

<i>Fold</i>	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
1	68,7	63,5	60,3
2	69,4	63,6	60,2
3	69,2	64,0	60,8
4	69,9	64,3	61,3
5	68,7	63,3	59,8
6	69,4	63,9	60,5
7	68,2	62,8	59,0
8	68,6	62,8	59,2
9	69,2	63,4	59,6
10	68,9	63,4	59,9
Rata-rata	69,0	63,5	60,0

Hasil dari Tabel 4.19 diketahui bahwa dari 10 percobaan yang dilakukan pada data *testing*, nilai rata-rata akurasi yang diperoleh menggunakan kernel linier hanya sebesar 69% dengan nilai AUC 63,5% dan nilai *G-Mean* sebesar 60%. Performa klasifikasi menggunakan SVM kernel linier masih memberikan hasil yang kurang baik untuk mengklasifikasikan pelanggan subsidi daya listrik 450 VA dan 900 VA.

Selain menggunakan kernel linier juga dilakukan klasifikasi SVM menggunakan kernel RBF. *Range* nilai parameter  $C$  dan  $\gamma$  menggunakan nilai  $C = 0,1, 1, 10$ , dan  $100$  dan nilai  $\gamma = 0,1, 1, 10$ , dan  $100$ . Klasifikasi SVM kernel RBF pada masing-masing *fold* data *training* dilakukan sebanyak 16 kombinasi  $C$  dan  $\gamma$ , sehingga total terdapat 160 kombinasi. Parameter  $C$  dan  $\gamma$  yang optimal yaitu parameter yang memberikan nilai rata-rata akurasi, AUC, dan *G-Mean* yang paling tinggi dari 10 percobaan data *training* (Huang & Wang, 2006). Selanjutnya, model SVM dengan parameter optimal tersebut akan diterapkan pada 10 data *testing* pada Tabel 4.21.

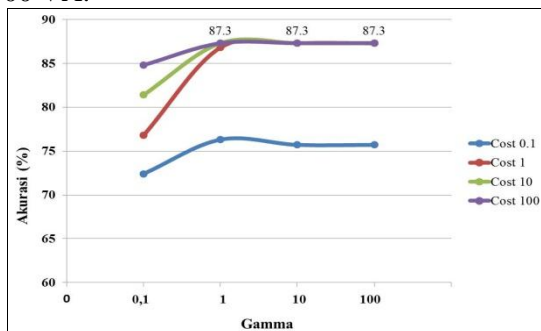
Hasil klasifikasi kernel RBF pada Tabel 4.20 dapat diketahui bahwa untuk mendapatkan hasil klasifikasi yang baik nilai parameter  $C$  dan  $\gamma$  yang harus di gunakan adalah  $C = 10$  dan  $100$  pada  $\gamma = 1$ . Selanjutnya dipilih nilai  $C = 10$  dan  $\gamma = 1$  karena rata-

rata nilai akurasi, AUC, dan *G-Mean* pada kombinasi tersebut telah memberikan hasil yang optimal.

**Tabel 4.20** Rata-rata Akurasi, AUC, dan *G-Mean* Kernel RBF (*Training*)

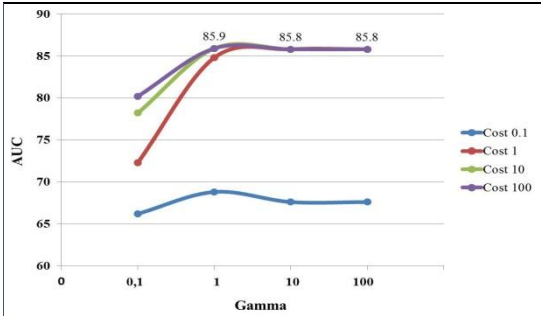
Cost (C)	Gamma ( $\gamma$ )	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
0,1	0,1	72,4	66,2	62,2
0,1	1	76,3	68,8	62,8
0,1	10	75,7	67,6	60,7
0,1	100	75,7	67,6	60,7
1	0,1	76,8	72,3	70,3
1	1	86,8	84,8	84,5
1	10	87,3	85,8	85,6
1	100	87,3	85,8	85,6
10	0,1	81,4	78,2	77,3
<b>10</b>	<b>1</b>	<b>87,3</b>	<b>85,9</b>	<b>85,7</b>
10	10	87,3	85,8	85,6
10	100	87,3	85,8	85,6
100	0,1	84,8	80,2	78,9
<b>100</b>	<b>1</b>	<b>87,3</b>	<b>85,9</b>	<b>85,7</b>
100	10	87,3	85,8	85,6
100	100	87,3	85,8	85,6

Selanjutnya secara visual akan dilihat pola klasifikasi yang diberikan oleh parameter *C* dan  $\gamma$  terhadap hasil akurasi, AUC, dan *G-Mean* pada klasifikasi pelanggan subsidi listrik daya 450 VA dan 900 VA.



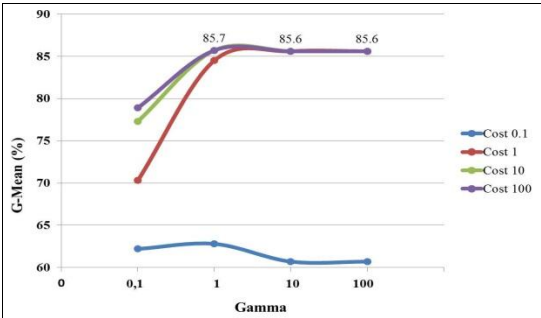
**Gambar 4.1** Pengaruh Nilai *Cost* dan *Gamma* pada Hasil Akurasi

Berdasarkan Gambar 4.1 dapat diketahui bahwa nilai *Gamma* ( $\gamma$ ) yang rendah yaitu pada  $\gamma = 0,1$  juga memberikan hasil akurasi yang rendah. Semakin tinggi nilai parameter *Cost* (*C*) yang digunakan akan menghasilkan nilai AUC yang semakin tinggi. Kombinasi nilai *C* = 10, 100 dengan nilai  $\gamma = 1, 10, 100$  memberikan tingkat akurasi dengan nilai yang sudah konstan.



**Gambar 4.2** Pengaruh Nilai *Cost* dan *Gamma* pada Hasil AUC

Berdasarkan Gambar 4.2 dapat diketahui bahwa semakin tinggi nilai parameter *Cost* (*C*) akan menghasilkan nilai AUC yang semakin tinggi, sedangkan untuk parameter *Gamma* ( $\gamma$ ) menghasilkan nilai AUC tertinggi pada saat  $\gamma = 1$  dan selanjutnya mengalami penurunan nilai AUC pada  $\gamma = 10$  dan 100 dengan nilai AUC yang sudah cenderung konstan. Pada saat  $\gamma = 0,1$  menghasilkan nilai AUC yang paling rendah.



**Gambar 4.3** Pengaruh Nilai *Cost* dan *Gamma* pada Hasil *G-Mean*



Berdasarkan Gambar 4.3, klasifikasi yang diberikan oleh parameter  $C$  dan  $\gamma$  terhadap nilai  $G$ -Mean memberikan pola klasifikasi yang hampir sama dengan nilai AUC. Semakin tinggi nilai parameter  $Cost$  ( $C$ ) akan menghasilkan nilai  $G$ -Mean yang semakin tinggi, sedangkan parameter gamma ( $\gamma$ ) menghasilkan nilai  $G$ -Mean tertinggi pada saat  $\gamma=1$ , mengalami penurunan nilai AUC pada  $\gamma=10$  dan  $\gamma=100$  dengan nilai  $G$ -Mean yang sudah cenderung konstan dan menghasilkan nilai  $G$ -Mean yang rendah pada saat  $\gamma=0,1$ .

Parameter optimal pada Tabel 4.20 selanjutnya diterapkan pada data *testing*, maka akan diperoleh performa klasifikasi berdasarkan rata-rata dari nilai akurasi, AUC, dan  $G$ -Mean (Huang & Wang, 2006).

**Tabel 4.21** Nilai Akurasi, AUC, dan  $G$ -Mean Kernel RBF (*Testing*)

<i>Fold</i>	Akurasi (%)	AUC (%)	$G$ -Means (%)
1	87,0	85,6	85,4
2	87,6	86,3	86,1
3	87,0	85,7	85,6
4	88,1	86,8	86,6
5	87,9	86,6	86,4
6	86,9	85,4	85,2
7	87,7	86,4	86,2
8	87,3	86,0	85,9
9	87,3	86,0	85,9
10	79,5	77,2	76,7
Rata-rata	86,6	85,2	85,0

Hasil dari Tabel 4.21 diketahui bahwa dari 10 percobaan yang dilakukan pada data *testing*, rata-rata nilai akurasi yang diperoleh sebesar 86,6% dengan nilai AUC 85,2% dan nilai  $G$ -Mean sebesar 85%. Performa klasifikasi menggunakan SVM kernel RBF sudah memberikan hasil yang cukup baik.

#### **4.2.2 Klasifikasi *Support Vector Machine* Menggunakan Variabel Hasil Seleksi FCBF**

Tahapan selanjutnya adalah melakukan klasifikasi dengan menggunakan variabel hasil seleksi FCBF yaitu variabel status

kepemilikan bangunan tempat tinggal, luas lantai, jenis lantai terluas, jumlah kamar tidur, kepemilikan lemari es, kepemilikan sepeda, dan status kesejahteraan.

**Tabel 4.22** Rata-rata Akurasi, AUC, dan *G-Mean* Kernel Linier (*Training*)

Cost ( <i>C</i> )	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
0,01	68,7	63,3	60,0
<b>0,1</b>	<b>68,7</b>	<b>63,3</b>	<b>60,0</b>
<b>1</b>	68,7	63,3	60,0
10	68,7	63,3	60,0
100	68,6	63,3	60,0

Hasil dari tabel 4.22 dapat diketahui bahwa rata-rata nilai akurasi, AUC, dan *G-Mean* pada semua nilai *C* menunjukkan nilai yang sudah konvergen. Selanjutnya dipilih nilai *C* = 0,1 sebagai parameter optimal untuk digunakan pada 10 data *testing*, sehingga akan diperoleh performa klasifikasi berdasarkan rata-rata dari nilai akurasi, AUC, dan *G-Mean* (Huang & Wang, 2006).

**Tabel 4.23** Nilai Akurasi, AUC, dan *G-Mean* Kernel Linier (*Testing*)

<i>Fold</i>	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
1	68,3	63,2	60,1
2	69,2	63,6	60,3
3	69,5	64,2	61,4
4	68,8	63,8	60,7
5	68,1	62,8	59,5
6	69,1	63,7	60,3
7	68,2	62,8	59,1
8	68,3	62,7	59,3
9	68,9	63,2	59,6
10	68,8	63,3	59,8
Rata-rata	68,7	63,3	60,0

Hasil dari Tabel 4.23 diketahui bahwa dari 10 percobaan yang dilakukan pada data *testing*, rata-rata nilai akurasi yang diperoleh hanya sebesar 68,7% dengan nilai AUC 63,3% dan *G-Mean* 60%. Hal ini menunjukkan bahwa performa klasifikasi dengan SVM kernel linier setelah dilakukan seleksi variabel

menghasilkan nilai akurasi, AUC, dan *G-Mean* yang tidak lebih baik dibandingkan menggunakan seluruh variabel prediktor.

Selain menggunakan kernel linier juga dilakukan klasifikasi SVM menggunakan kernel RBF. Pada masing-masing *fold* dilakukan sebanyak 16 kombinasi  $C$  dan  $\gamma$ , sehingga total terdapat 160 kombinasi. Tabel 4.24 merupakan nilai rata-rata akurasi, AUC, dan *G-Mean* untuk menentukan parameter yang optimal dengan memilih nilai performa yang paling tinggi.

**Tabel 4.24** Rata-rata Akurasi, AUC, dan *G-Mean* Kernel RBF (*Training*)

Cost ( $C$ )	Gamma ( $\gamma$ )	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
0,1	0,1	69,0	62,24	57,02
0,1	1	71,0	65,12	61,37
0,1	10	71,1	65,30	61,60
0,1	100	71,1	65,30	61,60
1	0,1	69,7	63,80	59,88
1	1	71,6	66,51	63,82
1	10	71,6	66,58	63,93
1	100	71,6	66,58	63,93
10	0,1	70,3	64,82	61,51
<b>10</b>	<b>1</b>	<b>71,6</b>	<b>66,61</b>	<b>64,01</b>
10	10	71,6	66,58	63,93
10	100	71,6	66,58	63,93
100	0,1	70,7	45,80	43,67
<b>100</b>	<b>1</b>	<b>71,6</b>	<b>66,61</b>	<b>64,01</b>
100	10	71,6	66,58	63,93
100	100	71,6	66,58	63,93

Hasil Tabel 4.24 dapat diketahui bahwa untuk mendapatkan hasil klasifikasi yang baik nilai parameter  $C$  dan  $\gamma$  yang harus digunakan adalah  $C = 10$  dan  $100$  pada  $\gamma = 1$ . Selanjutnya dipilih nilai  $C = 10$  dan  $\gamma = 1$  karena rata-rata hasil akurasi, AUC, dan *G-Mean* pada kombinasi tersebut telah memberikan hasil yang optimal. Parameter optimal  $C = 10$  dan  $\gamma = 1$  diterapkan pada 10 *fold* data *testing* pada Tabel 4.25.

Hasil dari Tabel 4.25 diketahui bahwa dari 10 percobaan yang dilakukan pada data *testing*, rata-rata nilai akurasi yang diperoleh sebesar 71,5% dengan nilai AUC 66,8% dan nilai *G-*

*Mean* sebesar 64,4%. Performa klasifikasi SVM kernel RBF setelah dilakukan seleksi variabel menghasilkan nilai akurasi, AUC, dan *G-Mean* yang tidak lebih baik apabila dibandingkan menggunakan seluruh variabel prediktor.

**Tabel 4.25** Nilai Akurasi, AUC, dan *G-Mean* Kernel RBF (*Testing*)

<i>Fold</i>	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
1	69,9	65,4	63,1
2	72,4	67,2	64,6
3	71,6	67,3	65,2
4	72,2	67,5	65,5
5	71,4	66,7	64,4
6	71,3	66,4	63,9
7	71,2	66,4	63,7
8	71,3	66,4	63,8
9	71,9	67,0	64,5
10	72,0	67,4	65,0
Rata-rata	71,5	66,8	64,4

#### 4.4 Klasifikasi Menggunakan Metode *Naïve Bayes Classifier*

Klasifikasi pada metode NBC akan dilakukan 2 tahapan yaitu klasifikasi menggunakan seluruh variabel prediktor, serta klasifikasi dengan variabel hasil seleksi FCBF. Partisi data yang digunakan sama dengan metode SVM yaitu membagi data *training* dan data *testing* menjadi 10 *fold cross validation*.

**Tabel 4.26** Hasil Klasifikasi *Naïve Bayes Classifier* Tanpa Seleksi Variabel

<i>Fold</i>	Akurasi	AUC	<i>G-Mean</i>
1	66,9	64,5	63,8
2	67,9	65,3	64,6
3	67,8	65,5	64,9
4	67,9	65,6	65,1
5	66,8	64,4	63,8
6	67,3	64,6	63,8
7	67,0	64,5	63,7
8	66,6	63,8	62,9
9	67,5	64,9	64,2
10	67,3	64,9	64,2
Rata-rata	67,3	64,8	64,1

Hasil dari Tabel 4.26 diketahui bahwa dari 10 percobaan yang dilakukan pada data *testing*, performa klasifikasi berdasarkan rata-rata nilai akurasi diperoleh sebesar 67,3% dengan nilai AUC 64,8% dan nilai *G-Mean* 64,1% (Huang & Wang, 2006). Performa klasifikasi *Naïve Bayes Classifier* berdasarkan nilai akurasi, AUC, dan *G-Mean* memberikan hasil yang rendah pada kasus klasifikasi pelanggan subsidi listrik rumah tangga.

**Tabel 4.27** Hasil Klasifikasi *Naïve Bayes Classifier* Seleksi Variabel FCBF

<i>Fold</i>	Akurasi	AUC	<i>G-Mean</i>
1	68.8	64,1	63,3
2	68.5	65,4	64,5
3	67.6	65,1	64,4
4	67.3	64,7	64,0
5	67,1	64,6	63,9
6	67,6	64,6	63,6
7	67,0	64,1	63,1
8	66,6	63,5	62,4
9	67,7	64,6	63,7
10	67,5	64,7	63,8
Rata-rata	67,6	64,5	63,7

Hasil dari Tabel 4.27 diketahui bahwa dari 10 percobaan yang dilakukan pada data *testing*, rata-rata akurasi yang diperoleh adalah 67,6% dengan nilai AUC 64,5% dan nilai *G-Mean* sebesar 63,7%. Performa klasifikasi *Naïve Bayes Classifier* setelah dilakukan seleksi variabel memberikan hasil akurasi, AUC, dan *G-Mean* lebih rendah dibandingkan menggunakan seluruh variabel prediktor. Selanjutnya dilakukan perhitungan data *testing* yang pertama menggunakan model *Naïve Bayes* dengan seluruh variabel prediktor.

**Tabel 4.28** Peluang Prior

Kategori	Peluang
Subsidi 450 VA	0,634
Subsidi 900 VA	0,366

Peluang masing-masing kategori pada setiap variabel prediktor ditunjukkan pada Tabel 4.29, 4.30, dan 4.31.

**Tabel 4.29** Peluang Kategori Variabel Prediktor

$X_1$	$P(K_1   450 \text{ VA}) = 0,306$	$P(K_1   900 \text{ VA}) = 0,303$
	$P(K_2   450 \text{ VA}) = 0,478$	$P(K_2   900 \text{ VA}) = 0,470$
	$P(K_3   450 \text{ VA}) = 0,187$	$P(K_3   900 \text{ VA}) = 0,196$
	$P(K_4   450 \text{ VA}) = 0,029$	$P(K_4   900 \text{ VA}) = 0,031$
$X_2$	$P(K_1   450 \text{ VA}) = 0,175$	$P(K_1   900 \text{ VA}) = 0,100$
	$P(K_2   450 \text{ VA}) = 0,273$	$P(K_2   900 \text{ VA}) = 0,156$
	$P(K_3   450 \text{ VA}) = 0,552$	$P(K_3   900 \text{ VA}) = 0,744$
$X_3$	$P(K_1   450 \text{ VA}) = 0,658$	$P(K_1   900 \text{ VA}) = 0,473$
	$P(K_2   450 \text{ VA}) = 0,239$	$P(K_2   900 \text{ VA}) = 0,356$
	$P(K_3   450 \text{ VA}) = 0,103$	$P(K_3   900 \text{ VA}) = 0,171$
$X_4$	$P(K_1   450 \text{ VA}) = 0,015$	$P(K_1   900 \text{ VA}) = 0,008$
	$P(K_2   450 \text{ VA}) = 0,222$	$P(K_2   900 \text{ VA}) = 0,144$
	$P(K_3   450 \text{ VA}) = 0,311$	$P(K_3   900 \text{ VA}) = 0,227$
	$P(K_4   450 \text{ VA}) = 0,448$	$P(K_4   900 \text{ VA}) = 0,616$
	$P(K_5   450 \text{ VA}) = 0,004$	$P(K_5   900 \text{ VA}) = 0,005$
$X_5$	$P(K_1   450 \text{ VA}) = 0,038$	$P(K_1   900 \text{ VA}) = 0,034$
	$P(K_2   450 \text{ VA}) = 0,100$	$P(K_2   900 \text{ VA}) = 0,048$
	$P(K_3   450 \text{ VA}) = 0,862$	$P(K_3   900 \text{ VA}) = 0,918$
$X_6$	$P(K_1   450 \text{ VA}) = 0,039$	$P(K_1   900 \text{ VA}) = 0,026$
	$P(K_2   450 \text{ VA}) = 0,261$	$P(K_2   900 \text{ VA}) = 0,270$
	$P(K_3   450 \text{ VA}) = 0,674$	$P(K_3   900 \text{ VA}) = 0,670$
	$P(K_4   450 \text{ VA}) = 0,005$	$P(K_4   900 \text{ VA}) = 0,011$
	$P(K_5   450 \text{ VA}) = 0,005$	$P(K_5   900 \text{ VA}) = 0,004$
	$P(K_6   450 \text{ VA}) = 0,017$	$P(K_6   900 \text{ VA}) = 0,019$

**Tabel 4.30** Peluang Kategori Variabel Prediktor (Lanjutan)

$X_7$	$P(K_1 450 \text{ VA}) = 0,642$	$P(K_1 900 \text{ VA}) = 0,400$
	$P(K_2 450 \text{ VA}) = 0,347$	$P(K_2 900 \text{ VA}) = 0,574$
	$P(K_3 450 \text{ VA}) = 0,011$	$P(K_3 900 \text{ VA}) = 0,026$
$X_8$	$P(K_1 450 \text{ VA}) = 0,018$	$P(K_1 900 \text{ VA}) = 0,013$
	$P(K_2 450 \text{ VA}) = 0,006$	$P(K_2 900 \text{ VA}) = 0,004$
	$P(K_3 450 \text{ VA}) = 0,176$	$P(K_3 900 \text{ VA}) = 0,135$
	$P(K_4 450 \text{ VA}) = 0,289$	$P(K_4 900 \text{ VA}) = 0,285$
	$P(K_5 450 \text{ VA}) = 0,480$	$P(K_5 900 \text{ VA}) = 0,483$
	$P(K_6 450 \text{ VA}) = 0,032$	$P(K_6 900 \text{ VA}) = 0,080$
$X_9$	$P(K_1 450 \text{ VA}) = 0,006$	$P(K_1 900 \text{ VA}) = 0,009$
	$P(K_2 450 \text{ VA}) = 0,029$	$P(K_2 900 \text{ VA}) = 0,014$
	$P(K_3 450 \text{ VA}) = 0,005$	$P(K_3 900 \text{ VA}) = 0,007$
	$P(K_4 450 \text{ VA}) = 0,915$	$P(K_4 900 \text{ VA}) = 0,939$
	$P(K_5 450 \text{ VA}) = 0,020$	$P(K_5 900 \text{ VA}) = 0,011$
	$P(K_6 450 \text{ VA}) = 0,026$	$P(K_6 900 \text{ VA}) = 0,020$
$X_{10}$	$P(K_1 450 \text{ VA}) = 0,679$	$P(K_1 900 \text{ VA}) = 0,477$
	$P(K_2 450 \text{ VA}) = 0,321$	$P(K_2 900 \text{ VA}) = 0,523$
$X_{11}$	$P(K_1 450 \text{ VA}) = 0,129$	$P(K_1 900 \text{ VA}) = 0,086$
	$P(K_2 450 \text{ VA}) = 0,871$	$P(K_2 900 \text{ VA}) = 0,914$
$X_{12}$	$P(K_1 450 \text{ VA}) = 0,979$	$P(K_1 900 \text{ VA}) = 0,944$
	$P(K_2 450 \text{ VA}) = 0,021$	$P(K_2 900 \text{ VA}) = 0,056$
$X_{13}$	$P(K_1 450 \text{ VA}) = 0,624$	$P(K_1 900 \text{ VA}) = 0,438$
	$P(K_2 450 \text{ VA}) = 0,376$	$P(K_2 900 \text{ VA}) = 0,562$
$X_{14}$	$P(K_1 450 \text{ VA}) = 0,445$	$P(K_1 900 \text{ VA}) = 0,359$
	$P(K_2 450 \text{ VA}) = 0,555$	$P(K_2 900 \text{ VA}) = 0,641$

**Tabel 4.31** Peluang Kategori Variabel Prediktor (Lanjutan)

$X_{15}$	$P(K_1   450 \text{ VA}) = 0,395$	$P(K_1   900 \text{ VA}) = 0,434$
	$P(K_2   450 \text{ VA}) = 0,604$	$P(K_2   900 \text{ VA}) = 0,566$
$X_{16}$	$P(K_1   450 \text{ VA}) = 0,244$	$P(K_1   900 \text{ VA}) = 0,117$
	$P(K_2   450 \text{ VA}) = 0,281$	$P(K_2   900 \text{ VA}) = 0,246$
	$P(K_3   450 \text{ VA}) = 0,256$	$P(K_3   900 \text{ VA}) = 0,290$
	$P(K_4   450 \text{ VA}) = 0,218$	$P(K_4   900 \text{ VA}) = 0,347$

Nilai peluang pada Tabel 4.29, 4.30, dan 4.31 digunakan untuk menentukan peluang parsial pada setiap data *testing*. Selanjutnya menghitung nilai peluang parsial pada data *testing* pertama ditunjukkan pada Tabel 4.32.

**Tabel 4.32** Peluang Parsial Data *Testing* Pertama *Fold 4*

	$x_{li}$	$P(x_{li}   450 \text{ VA})$	$P(x_{li}   900 \text{ VA})$
$X_1$	2	0,478	0,470
$X_2$	3	0,552	0,744
$X_3$	1	0,658	0,473
$X_4$	4	0,448	0,616
$X_5$	3	0,862	0,918
$X_6$	3	0,674	0,670
$X_7$	1	0,642	0,400
$X_8$	5	0,048	0,483
$X_9$	4	0,915	0,939
$X_{10}$	1	0,679	0,477
$X_{11}$	2	0,871	0,914
$X_{12}$	1	0,979	0,944
$X_{13}$	1	0,624	0,438
$X_{14}$	2	0,555	0,641
$X_{15}$	2	0,604	0,566
$X_{16}$	2	0,281	0,246

Setelah menemukan peluang parsial tiap variabel pada setiap kategori kelas pada Tabel 4.32, langkah selanjutnya adalah menghitung peluang posterior yang kemudian digunakan untuk menentukan kelas klasifikasi pada data *testing*.



$$P(x_1, x_2, \dots, x_{16} | 450 \text{ VA}) = P(x_{1,1} | 450 \text{ VA}) \times \dots \times P(x_{1,16} | 450 \text{ VA}) \\ = 4.33638\text{e-}05$$

$$P(x_1, x_2, \dots, x_{16} | 900 \text{ VA}) = P(x_{1,1} | 900 \text{ VA}) \times \dots \times P(x_{1,16} | 900 \text{ VA}) \\ = 0.000183$$

Hasil perhitungan peluang posterior pada data *testing* pertama menunjukkan bahwa peluang terbesar adalah kategori subsidi 900 VA, sehingga data *testing* tersebut akan diklasifikasikan ke dalam pelanggan rumah tangga subsidi 900 VA.

#### 4.4 Perbandingan Metode SVM dan NBC

Langkah selanjutnya adalah membandingkan performa yang dihasilkan dari metode SVM dan NBC dengan membandingkan nilai akurasi, AUC, dan *G-Mean*.

**Tabel 4.33** Perbandingan Akurasi

Metode	Akurasi (%)	AUC (%)	<i>G-Mean</i> (%)
SVM Linier	69,0	63,5	60,0
<b>SVM RBF</b>	<b>86,6</b>	<b>85,2</b>	<b>85,0</b>
SVM Linier Seleksi	68,7	63,3	60,0
SVM RBF Seleksi	71,5	66,8	64,4
NBC	67,3	64,8	64,1
NBC seleksi	67,6	64,5	63,7

Tabel 4.33 dapat dilihat bahwa dari hasil pengukuran akurasi, AUC, dan *G-Mean* untuk metode SVM kernel RBF lebih baik dibandingkan metode SVM kernel linier, serta hasil klasifikasi menggunakan seluruh variabel prediktor lebih baik dibandingkan menggunakan hasil seleksi variabel FCBF. Hasil klasifikasi menggunakan metode NBC juga diperoleh performa klasifikasi lebih tinggi dengan menggunakan seluruh variabel prediktor dibandingkan menggunakan prediktor hasil seleksi FCBF. Perbandingan antara metode SVM dan NBC diperoleh hasil bahwa metode SVM lebih baik dibandingkan metode NBC. Nilai akurasi, AUC, dan *G-Mean* untuk metode SVM kernel RBF tanpa seleksi sudah mencapai performa diatas 80% dibandingkan metode lainnya yang hanya memberikan hasil performansi rata-rata 60%.

Berdasarkan Tabel 4.33 memberikan hasil bahwa metode SVM kernel RBF tanpa seleksi variabel memberikan performa yang paling baik. Berdasarkan nilai parameter optimal yang telah diperoleh pada tabel 4.20 sebesar  $C = 10$  dan  $\gamma = 1$ , fungsi *hyperplane* yang terbentuk untuk klasifikasi pelanggan rumah tangga subsidi 450 VA dan pelanggan subsidi 900 VA adalah:

$$D(x) = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b$$

dimana fungsi kernel yang digunakan adalah *Radial Basis Function* (RBF) dengan parameter  $\gamma = 1$ , yaitu dengan rumus:

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}) &= \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \\ &= \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \end{aligned}$$

diketahui  $\mathbf{x}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i16}]$  dengan  $x$  merupakan variabel prediktor sebanyak 16 variabel, indeks  $i$  sebanyak data *training* yaitu  $i = 1, 2, \dots, 71337$ , serta indeks  $j$  sebanyak data *testing* yang dicobakan. Diperoleh nilai  $\alpha_i$  berukuran  $71337 \times 1$  dan nilai  $b = -0,6973412$ , sehingga fungsi *hyperplane* yang diperoleh menjadi:

$$\begin{aligned} D(x) &= \sum_{i=1}^N \alpha_i y_i \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) + b \\ D(x) &= \sum_{i=1}^{71337} \alpha_i y_i \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) - 0,6973412. \end{aligned}$$

Hasil klasifikasi yang diperoleh apabila nilai  $D(x) < 0$  maka dikategorikan ke dalam pelanggan rumah tangga subsidi 450 VA, sedangkan apabila  $D(x) \geq 0$  maka dikategorikan pelanggan rumah tangga subsidi 900 VA.

Nilai parameter optimal yang telah diperoleh pada tabel 4.20 sebesar  $C = 10$  dan  $\gamma = 1$  pada kernel RBF selanjutnya digunakan untuk menghitung nilai *confusion matrix* menggunakan *fold* ke-4 hasil dari Tabel 4.21 dengan nilai akurasi, AUC, dan *G-Mean* yang tertinggi.

**Tabel 4.34** Hasil *Confusion Matrix*

Kelas aktual	Kelas prediksi				Jumlah
	450 VA	%	900 VA	%	
450 VA	4668	91,6	429	8,4	5097
900 VA	512	18,1	2318	81,9	2830

Tabel 4.34 menunjukkan bahwa kategori 450 VA sebanyak 91,6% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 8,4% pelanggan rumah tangga diprediksi masuk ke dalam kategori subsidi 900 VA. Pada kategori subsidi 900 VA sebanyak 81,9% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 18,1% pelanggan rumah tangga diprediksi masuk ke dalam kategori 450 VA.

Selanjutnya akan dilakukan karakteristik pelanggan rumah tangga dari hasil klasifikasi menggunakan SVM kernel RBF. Karakteristik pelanggan dilakukan berdasarkan variabel pembeda antara pelanggan subsidi 450 VA dengan pelanggan subsidi 900 VA yaitu jumlah kamar tidur, kepemilikan lemari es/kulkas, kepemilikan sepeda, dan status kesejahteraan. Variabel tersebut didapatkan dari karakteristik awal yang telah dilakukan pada tabel 4.1 hingga 4.16, serta hasil seleksi variabel menggunakan FCBF.

**Tabel 4.35** Karakteristik Jumlah Kamar Tidur

Kelas aktual	Kelas prediksi	%		
		0 – 1	2 – 3	>3
450 VA	450 VA	<b>66,195</b>	32,841	0,964
450 VA	900 VA	32,401	<b>66,900</b>	0,699
900 VA	900 VA	33,736	<b>63,028</b>	3,236
900 VA	450 VA	<b>61,133</b>	37,891	0,977

Hasil karakteristik berdasarkan Tabel 4.35 pelanggan rumah tangga subsidi 450 yang diprediksi menjadi pelanggan subsidi 900 VA sebagian besar memiliki jumlah kamar sebanyak 2

hingga 3 kamar, hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA. Sebanyak 429 rumah tangga dengan daya 450 VA disalahklasifikasikan menjadi pelanggan subsidi daya 900 VA. Pelanggan rumah tangga subsidi 900 sebanyak 512 pelanggan dianggap mendekati karakteristik pelanggan subsidi 450 VA yaitu hanya memiliki 1 kamar atau tidak memiliki kamar tidur di dalam tempat tinggalnya.

**Tabel 4.36** Karakteristik Kepemilikan Lemari Es/Kulkas

Kelas aktual	Kelas Prediksi	%	
		Tidak	Ya
450 VA	450 VA	<b>70,716</b>	29,284
450 VA	900 VA	43,357	<b>56,643</b>
900 VA	900 VA	41,588	<b>58,412</b>
900 VA	450 VA	<b>64,648</b>	35,352

Hasil karakteristik berdasarkan Tabel 4.36 pelanggan rumah tangga subsidi 450 yang diprediksi menjadi pelanggan subsidi 900 VA sebesar 56,643% sudah memiliki lemari es/kulkas. Hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA yang sebagian besar pelanggannya sudah memiliki lemari es/kulkas. Pelanggan rumah tangga subsidi 900 sebanyak 512 pelanggan sebagian besar dianggap mendekati karakteristik pelanggan subsidi 450 VA yaitu 64,648% masih tidak memiliki lemari es/kulkas.

**Tabel 4.37** Karakteristik Kepemilikan Sepeda

Kelas aktual	Kelas Prediksi	%	
		Tidak	Ya
450 VA	450 VA	<b>64,310</b>	35,690
450 VA	900 VA	46,620	<b>53,38</b>
900 VA	900 VA	38,179	<b>61,821</b>
900 VA	450 VA	<b>63,086</b>	36,914

Hasil karakteristik berdasarkan Tabel 4.37 pelanggan rumah tangga subsidi 450 yang salah diklasifikasikan menjadi pelanggan subsidi 900 VA dari total 429 sebanyak 53,38% memiliki sepeda. Hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA yang sebagian besar pelanggannya memiliki sepeda. Pelanggan rumah tangga subsidi daya 900 VA yang

disalahklasifikasikan menjadi pelanggan subsidi 450 VA yaitu sebesar 63,086% dari 512 pelanggan tidak memiliki sepeda.

**Tabel 4.38** Status Kesejahteraan

Kelas aktual	Kelas prediksi	%			
		Status 1	Status 2	Status 3	Status 4
450 VA	450 VA	25,386	<b>28,899</b>	25,643	20,073
450 VA	900 VA	7,226	22,844	31,702	<b>38,228</b>
900 VA	900 VA	11,001	23,425	28,775	<b>36,799</b>
900 VA	450 VA	14.648	27.148	<b>32.227</b>	25.977

Hasil karakteristik berdasarkan Tabel 4.38 pelanggan rumah tangga subsidi 450 yang salah diklasifikasikan menjadi pelanggan subsidi 900 VA sebesar 38,228% dari 429 pelanggan sudah memiliki status kesejahteraan 31% - 40% terendah, hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA. Pelanggan rumah tangga subsidi 900 VA yang salah di klasifikasikan menjadi pelanggan subsidi 450 VA sebesar 32,227% dari 512 pelanggan dengan status kesejahteraan 21% - 30% terendah, hal ini mendekati karakteristik yang dimiliki pelanggan subsidi 450 VA sebagian besar dengan status kesejahteraan 11% - 20%.

*(Halaman ini sengaja dikosongkan)*

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Hasil karakteristik dari pelanggan dan hasil seleksi variabel dengan metode FCBF diperoleh variabel yang membedakan antara karakteristik pelanggan subsidi 450 VA dan 900 VA adalah variabel jumlah kamar tidur, kepemilikan lemari es, kepemilikan sepeda, dan status kesejahteraan. Hasil pengukuran akurasi, AUC, dan *G-mean* untuk metode SVM kernel RBF lebih baik dibandingkan metode SVM kernel linier, serta hasil klasifikasi menggunakan seluruh variabel prediktor lebih baik dibandingkan menggunakan hasil seleksi variabel FCBF. Hasil klasifikasi menggunakan metode NBC juga diperoleh performa klasifikasi lebih tinggi dengan menggunakan seluruh variabel prediktor dibandingkan menggunakan prediktor hasil seleksi FCBF. Perbandingan antara metode SVM dan NBC diperoleh hasil bahwa metode SVM lebih baik dibandingkan metode NBC dengan parameter optimal yang diperoleh adalah  $C = 10$  dan  $\gamma = 1$ . Kategori 450 VA sebanyak 91,6% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 8,4% pelanggan rumah tangga diprediksi masuk ke dalam kategori subsidi 900 VA. Pada kategori subsidi 900 VA sebanyak 81,9% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 18,1% pelanggan rumah tangga diprediksi masuk ke dalam kategori 450 VA.

#### **5.2 Saran**

Saran yang dapat diberikan pada penelitian selanjutnya adalah lebih memperhatikan faktor atau variabel yang berpengaruh terhadap daya listrik rumah tangga yang memperoleh subsidi listrik, seperti faktor dari ketenagalistrikan yaitu jam nyala, pemakaian kWh, tarif listrik, dan lain sebagainya. Metode yang digunakan bisa dikembangkan dengan metode-

metode lain sebagai perbandingan terhadap metode *Support Vector Machine* dan *Naïve Bayes Classifier*. Perbandingan kernel pada klasifikasi SVM juga dapat digunakan selain kernel Linier dan RBF sebagai perbandingan untuk memperoleh hasil yang paling baik. Data yang digunakan untuk analisis klasifikasi lebih baik menggunakan data yang *balance* antar kategori kelasnya untuk meningkatkan performa klasifikasi.



## DAFTAR PUSTAKA

- Bekkar, M., Djema, H. K., & Alitouche, T. A. (2013). Evaluation Measures for Models Assesment over Imbalanced Datasets. *Journal of Information Engineering and Application*, 3(10), 27-39.
- Bengio, Y. (2004). No Unbiased Estimator of The Variance of K-Fold Cross Validation. *Journal of Machine Learning Research*, 5, 1089-1105.
- BPS Provinsi Jawa Timur. (2017). *Indikator Kesejahteraan Rakyat Jawa Timur 2017*. Surabaya: BPS Provinsi Jawa Timur.
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Journal of Machine Learning Research*, 20(3), 273-297.
- Ditjen Ketenagalistrikan KESDM. (2016). *Statistik Ketenagalistrikan 2017*. Jakarta: Ditjen Ketenagalistrikan KESDM.
- Gokgoz, E., & Subasi, A. (2015). Comparison of decision tree algorithms for EMG signal classification using DWT. *Journal of Biomedical Signal Processing and Control*, 18, 138–144.
- Gunn, S. (1998). *Support vector Machines fo Classification and Regression*. Southampton, UK: University of Southampton.
- Han, J. & Kamber, M., (2006). *Data Mining: Concepts and Techniques*, 2<sup>nd</sup> Edition. San Francisco: Morgan Kaufmann.
- Hsu, C., Chang, C., & Lin, C. (2004). *A Practical Guide to Support Vector Classification*. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University.
- Huang, C. L., & Wang, C. J. (2006). A GA-Based Feature Selection and Parameters Optimization for Support Vector Machines. *International Journal of Expert Systems with Applications*, 31(2), 231-240.
- Kohavi R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the*

- 14<sup>th</sup> *International Joint Conference on Artificial Intelligence*, 2, 1137-1143.
- Larose DT. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Liu, H., & Lei, Y. (2003). Feature Selection for High Dimensional Data: Fast Correlation Based Filter Solution. *Proceeding of Twentieth International Conference on Machine Learning*, 2, 856-863.
- PLN. (2017). Laporan Manajemen Perusahaan. Jawa Timur, Surabaya: PT. PLN (Persero) Distribusi Jawa Timur.
- Sahitayakti, R. P., & Fithriasari, K. (2015). Klasifikasi Kesejahteraan Rumah Tangga di Provinsi Papua dengan Metode Regresi Logistik dan *Support Vector Machine*. *Jurnal Sains dan Seni ITS*, 4(2), 2337-3520.
- Sain, H., & Purnami, S. W. (2015). Combine Sampling Support Vactor Machine for Imbalanced Data Classification. *Procedia Computer Science*, 72, 59 - 66.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, 2(3).
- Scholkopf, B., & Smola, A. (2002). *Learning with Kernel: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: MIT Press.
- Siregar, S. A., & Warman, E. (2013). Studi Prakiraan Kebutuhan Energi Listrik Tahun 2013-2017 Wilayah Kota Padang Sidempuan dengan Metode Gabungan. *Jurnal Singuda Ensikom*, 1(2).
- Taheri, S., & Mammadov, M. (2013). Learning The Naïve Bayes Classifier with Optimization Models, *International Journal of Applied Mathematics and Computer Science*, 23(4).
- Tanjung, Y. P., Sentinuwo, S., & Jacobus, A. (2016). Penentuan Daya Listrik Rumah Tangga Menggunakan Metode *Decision Tree*. *Jurnal Teknik Informatika*, 9(1).

- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*, 3<sup>rd</sup> edition. San Francisco: Morgan Kaufmann Publishers.
- Wu, Xindong & Kumar, V. (2009). The Top Ten Algorithms in Data Mining. *Journal of Knowledge and Information Systems*, 14(1), 1-37.
- Zhang, H. (2004). The Optimality of Naïve Bayes. *Proceeding of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, 562-567.
- Nayak, A., & Natarajan, D. S. (2016). Comparative Study of Naïve Bayes Classifier in Sentiment Analysis of Twitter Feeds. *International Journal of Advance Studies in Computer Science and Engineering*, 5(1).

*(Halaman ini sengaja dikosongkan)*

# LAMPIRAN

## Lampiran 1. *Syntax Metode Support Vector Machine*

```
trainData <- read.csv(file.choose(),header=TRUE)
testData <- read.csv(file.choose(),header=TRUE)
train_labels <- as.factor(trainData$Daya.listrik)
test_labels <- as.factor(testData$Daya.listrik)

# Klasifikasi kernel RBF
svm_model<- svm(as.factor(Daya.listrik) ~ ., data=trainData,
                kernel="radial", cost=10, gamma=1)

# Klasifikasi kernel Linier
svm_model<- svm(as.factor(Daya.listrik) ~ ., data=trainData,
                kernel="linear", cost=0,1)

# Menentukan Parameter Optimal Menggunakan Data Training
pred <- predict(svm_model,trainData)
CrossTable(x=train_labels,y=pred, prop.chisq=FALSE)

# Menentukan Tingkat Akurasi Pada Data Testing
pred <- predict(svm_model,testData)
CrossTable(x=test_labels,y=pred, prop.chisq=FALSE)
```

## Lampiran 2. *Syntax Metode Naïve Bayes Classifier*

```
trainData <- read.csv(file.choose(),header=TRUE)
testData <- read.csv(file.choose(),header=TRUE)
train_labels <- as.factor(trainData$Daya.listrik)
test_labels <- as.factor(testData$Daya.listrik)
trainData[,1]<- as.factor(trainData[,1])
trainData[,2]<- as.factor(trainData[,2])
trainData[,3]<- as.factor(trainData[,3])
trainData[,4]<- as.factor(trainData[,4])
trainData[,5]<- as.factor(trainData[,5])
```

```
trainData[,6]<- as.factor(trainData[,6])
trainData[,7]<- as.factor(trainData[,7])
trainData[,8]<- as.factor(trainData[,8])
trainData[,9]<- as.factor(trainData[,9])
trainData[,10]<- as.factor(trainData[,10])
trainData[,11]<- as.factor(trainData[,11])
trainData[,12]<- as.factor(trainData[,12])
trainData[,13]<- as.factor(trainData[,13])
trainData[,14]<- as.factor(trainData[,14])
trainData[,15]<- as.factor(trainData[,15])
trainData[,16]<- as.factor(trainData[,16])
trainData[,17]<- as.factor(trainData[,17])
testData[,1]<- as.factor(testData[,1])
testData[,2]<- as.factor(testData[,2])
testData[,3]<- as.factor(testData[,3])
testData[,4]<- as.factor(testData[,4])
testData[,5]<- as.factor(testData[,5])
testData[,6]<- as.factor(testData[,6])
testData[,7]<- as.factor(testData[,7])
testData[,8]<- as.factor(testData[,8])
testData[,9]<- as.factor(testData[,9])
testData[,10]<- as.factor(testData[,10])
testData[,11]<- as.factor(testData[,11])
testData[,12]<- as.factor(testData[,12])
testData[,13]<- as.factor(testData[,13])
testData[,14]<- as.factor(testData[,14])
testData[,15]<- as.factor(testData[,15])
testData[,16]<- as.factor(testData[,16])
testData[,17]<- as.factor(testData[,17])
model <- naiveBayes(Daya.listrik ~ ., data = trainData)
pred <- predict(model,testData)
CrossTable(x=test_labels,y=pred, prop.chisq=FALSE)
```

### Lampiran 3. Output *Support Vector Machine* Kernel RBF

FOLD 1						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.725	0.891	0.435	0.663	0.623
0.1	1	0.764	0.967	0.41	0.689	0.630
0.1	10	0.757	0.975	0.377	0.676	0.606
0.1	100	0.757	0.975	0.377	0.676	0.606
1	0.1	0.768	0.892	0.552	0.722	0.702
1	1	0.868	0.921	0.776	0.849	0.845
1	10	0.874	0.918	0.797	0.858	0.855
1	100	0.874	0.918	0.797	0.858	0.855
10	0.1	0.815	0.899	0.669	0.784	0.776
10	1	0.874	0.914	0.803	0.859	0.857
10	10	0.874	0.918	0.797	0.858	0.855
10	100	0.874	0.918	0.797	0.858	0.855
100	0.1	0.849	0.908	0.746	0.827	0.823
100	1	0.874	0.914	0.803	0.859	0.857
100	10	0.874	0.918	0.797	0.858	0.855
100	100	0.874	0.918	0.797	0.858	0.855

FOLD 2						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.722	0.888	0.435	0.662	0.622
0.1	1	0.763	0.967	0.41	0.689	0.630
0.1	10	0.757	0.974	0.38	0.677	0.608
0.1	100	0.757	0.974	0.38	0.677	0.608
1	0.1	0.768	0.889	0.557	0.723	0.704
1	1	0.867	0.919	0.776	0.848	0.844
1	10	0.872	0.916	0.798	0.857	0.855
1	100	0.872	0.916	0.798	0.857	0.855
10	0.1	0.814	0.9	0.665	0.783	0.774
10	1	0.873	0.913	0.804	0.859	0.857
10	10	0.872	0.916	0.798	0.857	0.855
10	100	0.872	0.916	0.798	0.857	0.855
100	0.1	0.848	0.907	0.747	0.827	0.823
100	1	0.873	0.913	0.804	0.859	0.857
100	10	0.872	0.916	0.798	0.857	0.855
100	100	0.872	0.916	0.798	0.857	0.855

FOLD 3						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.723	0.893	0.428	0.661	0.618
0.1	1	0.763	0.967	0.41	0.689	0.630
0.1	10	0.758	0.975	0.38	0.678	0.609
0.1	100	0.758	0.975	0.38	0.678	0.609
1	0.1	0.768	0.892	0.553	0.723	0.702
1	1	0.867	0.92	0.777	0.849	0.845
1	10	0.874	0.917	0.797	0.857	0.855
1	100	0.874	0.917	0.797	0.857	0.855
10	0.1	0.813	0.903	0.656	0.780	0.770
10	1	0.874	0.914	0.803	0.859	0.857
10	10	0.874	0.917	0.797	0.857	0.855
10	100	0.874	0.917	0.797	0.857	0.855
100	0.1	0.847	0.909	0.27	0.590	0.495
100	1	0.874	0.914	0.803	0.859	0.857
100	10	0.874	0.917	0.797	0.857	0.855
100	100	0.874	0.917	0.797	0.857	0.855

FOLD 4						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.724	0.889	0.437	0.663	0.623
0.1	1	0.762	0.968	0.407	0.688	0.628
0.1	10	0.757	0.975	0.379	0.677	0.608
0.1	100	0.757	0.975	0.379	0.677	0.608
1	0.1	0.767	0.891	0.553	0.722	0.702
1	1	0.866	0.919	0.775	0.847	0.844
1	10	0.872	0.913	0.801	0.857	0.855
1	100	0.872	0.913	0.801	0.857	0.855
10	0.1	0.813	0.9	0.66	0.780	0.771
10	1	0.872	0.91	0.806	0.858	0.856
10	10	0.872	0.913	0.801	0.857	0.855
10	100	0.872	0.913	0.801	0.857	0.855
100	0.1	0.847	0.909	0.74	0.825	0.820
100	1	0.872	0.91	0.806	0.858	0.856
100	10	0.872	0.913	0.801	0.857	0.855
100	100	0.872	0.913	0.801	0.857	0.855



FOLD 5						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.723	0.89	0.434	0.662	0.621
0.1	1	0.763	0.968	0.406	0.687	0.627
0.1	10	0.756	0.977	0.371	0.674	0.602
0.1	100	0.756	0.977	0.371	0.674	0.602
1	0.1	0.767	0.892	0.552	0.722	0.702
1	1	0.868	0.918	0.779	0.849	0.846
1	10	0.873	0.914	0.802	0.858	0.856
1	100	0.873	0.914	0.802	0.858	0.856
10	0.1	0.813	0.901	0.661	0.781	0.772
10	1	0.873	0.911	0.806	0.859	0.857
10	10	0.873	0.914	0.802	0.858	0.856
10	100	0.873	0.914	0.802	0.858	0.856
100	0.1	0.848	0.909	0.743	0.826	0.822
100	1	0.873	0.911	0.806	0.859	0.857
100	10	0.873	0.914	0.802	0.858	0.856
100	100	0.873	0.914	0.802	0.858	0.856

FOLD 6						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.724	0.891	0.433	0.662	0.621
0.1	1	0.763	0.967	0.408	0.688	0.628
0.1	10	0.758	0.974	0.381	0.678	0.609
0.1	100	0.758	0.974	0.381	0.678	0.609
1	0.1	0.768	0.89	0.555	0.723	0.703
1	1	0.868	0.919	0.781	0.850	0.847
1	10	0.874	0.916	0.802	0.859	0.857
1	100	0.874	0.916	0.802	0.859	0.857
10	0.1	0.815	0.897	0.671	0.784	0.776
10	1	0.874	0.913	0.806	0.860	0.858
10	10	0.874	0.916	0.802	0.859	0.857
10	100	0.874	0.916	0.802	0.859	0.857
100	0.1	0.848	0.907	0.746	0.827	0.823
100	1	0.874	0.913	0.806	0.860	0.858
100	10	0.874	0.916	0.802	0.859	0.857
100	100	0.874	0.916	0.802	0.859	0.857

FOLD 7						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.725	0.893	0.434	0.664	0.623
0.1	1	0.763	0.968	0.405	0.687	0.626
0.1	10	0.758	0.975	0.378	0.677	0.607
0.1	100	0.758	0.975	0.378	0.677	0.607
1	0.1	0.767	0.889	0.555	0.722	0.702
1	1	0.867	0.921	0.774	0.848	0.844
1	10	0.873	0.915	0.8	0.858	0.856
1	100	0.873	0.915	0.8	0.858	0.856
10	0.1	0.814	0.9	0.663	0.782	0.772
10	1	0.872	0.912	0.805	0.859	0.857
10	10	0.873	0.915	0.8	0.858	0.856
10	100	0.873	0.915	0.8	0.858	0.856
100	0.1	0.847	0.911	0.736	0.824	0.819
100	1	0.872	0.912	0.805	0.859	0.857
100	10	0.873	0.915	0.8	0.858	0.856
100	100	0.873	0.915	0.8	0.858	0.856

FOLD 8						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.724	0.89	0.436	0.663	0.623
0.1	1	0.762	0.968	0.405	0.687	0.626
0.1	10	0.756	0.976	0.376	0.676	0.606
0.1	100	0.756	0.976	0.376	0.676	0.606
1	0.1	0.77	0.892	0.559	0.726	0.706
1	1	0.868	0.92	0.776	0.848	0.845
1	10	0.873	0.917	0.797	0.857	0.855
1	100	0.873	0.917	0.797	0.857	0.855
10	0.1	0.814	0.901	0.665	0.783	0.774
10	1	0.873	0.914	0.803	0.859	0.857
10	10	0.873	0.917	0.797	0.857	0.855
10	100	0.873	0.917	0.797	0.857	0.855
100	0.1	0.848	0.91	0.74	0.825	0.821
100	1	0.873	0.914	0.803	0.859	0.857
100	10	0.873	0.917	0.797	0.857	0.855
100	100	0.873	0.917	0.797	0.857	0.855

FOLD 9						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.725	0.892	0.434	0.663	0.622
0.1	1	0.763	0.968	0.408	0.688	0.628
0.1	10	0.755	0.976	0.373	0.675	0.603
0.1	100	0.755	0.976	0.373	0.675	0.603
1	0.1	0.767	0.892	0.549	0.721	0.700
1	1	0.868	0.922	0.774	0.848	0.845
1	10	0.873	0.916	0.8	0.858	0.856
1	100	0.873	0.916	0.8	0.858	0.856
10	0.1	0.814	0.901	0.663	0.782	0.773
10	1	0.874	0.914	0.804	0.859	0.857
10	10	0.873	0.916	0.8	0.858	0.856
10	100	0.873	0.916	0.8	0.858	0.856
100	0.1	0.849	0.91	0.741	0.826	0.821
100	1	0.874	0.914	0.804	0.859	0.857
100	10	0.873	0.916	0.8	0.858	0.856
100	100	0.873	0.916	0.8	0.858	0.856

FOLD 10						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.723	0.89	0.434	0.662	0.621
0.1	1	0.765	0.966	0.413	0.690	0.632
0.1	10	0.758	0.975	0.382	0.679	0.610
0.1	100	0.758	0.975	0.382	0.679	0.610
1	0.1	0.768	0.89	0.556	0.723	0.703
1	1	0.869	0.919	0.78	0.850	0.847
1	10	0.875	0.916	0.802	0.859	0.857
1	100	0.875	0.916	0.802	0.859	0.857
10	0.1	0.815	0.899	0.668	0.784	0.775
10	1	0.874	0.912	0.81	0.861	0.859
10	10	0.875	0.916	0.802	0.859	0.857
10	100	0.875	0.916	0.802	0.859	0.857
100	0.1	0.85	0.908	0.75	0.829	0.825
100	1	0.874	0.912	0.81	0.861	0.859
100	10	0.875	0.916	0.802	0.859	0.857
100	100	0.875	0.916	0.802	0.859	0.857

**Lampiran 4.** Output *Support Vector Machine* Kernel RBF  
Seleksi Variabel FCBF

FOLD 1						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.691	0.872	0.376	0.624	0.573
0.1	1	0.711	0.866	0.442	0.654	0.619
0.1	10	0.713	0.868	0.441	0.655	0.619
0.1	100	0.713	0.868	0.441	0.655	0.619
1	0.1	0.697	0.854	0.426	0.640	0.603
1	1	0.717	0.848	0.488	0.668	0.643
1	10	0.717	0.847	0.49	0.669	0.644
1	100	0.717	0.847	0.49	0.669	0.644
10	0.1	0.704	0.854	0.443	0.649	0.615
10	1	0.717	0.847	0.492	0.670	0.646
10	10	0.717	0.847	0.49	0.669	0.644
10	100	0.717	0.847	0.49	0.669	0.644
100	0.1	0.704	0.854	0.443	0.649	0.615
100	1	0.717	0.847	0.492	0.670	0.646
100	10	0.717	0.847	0.49	0.669	0.644
100	100	0.717	0.847	0.49	0.669	0.644

FOLD 2						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.689	0.871	0.373	0.622	0.570
0.1	1	0.709	0.87	0.431	0.651	0.612
0.1	10	0.71	0.87	0.434	0.652	0.614
0.1	100	0.71	0.87	0.434	0.652	0.614
1	0.1	0.696	0.86	0.413	0.637	0.596
1	1	0.715	0.851	0.478	0.665	0.638
1	10	0.715	0.851	0.48	0.666	0.639
1	100	0.715	0.851	0.48	0.666	0.639
10	0.1	0.702	0.853	0.442	0.648	0.614
10	1	0.715	0.85	0.481	0.666	0.639
10	10	0.715	0.851	0.48	0.666	0.639
10	100	0.715	0.851	0.48	0.666	0.639
100	0.1	0.707	0.846	0.466	0.656	0.628
100	1	0.715	0.85	0.481	0.666	0.639
100	10	0.715	0.851	0.48	0.666	0.639
100	100	0.715	0.851	0.48	0.666	0.639

FOLD 3						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.691	0.868	0.38	0.624	0.574
0.1	1	0.71	0.867	0.436	0.652	0.615
0.1	10	0.711	0.869	0.437	0.653	0.616
0.1	100	0.711	0.869	0.437	0.653	0.616
1	0.1	0.697	0.861	0.413	0.637	0.596
1	1	0.716	0.857	0.469	0.663	0.634
1	10	0.716	0.859	0.467	0.663	0.633
1	100	0.716	0.859	0.467	0.663	0.633
10	0.1	0.703	0.853	0.442	0.648	0.614
10	1	0.716	0.856	0.472	0.664	0.636
10	10	0.716	0.859	0.467	0.663	0.633
10	100	0.716	0.859	0.467	0.663	0.633
100	0.1	0.706	0.856	0.448	0.652	0.619
100	1	0.716	0.856	0.472	0.664	0.636
100	10	0.716	0.859	0.467	0.663	0.633
100	100	0.716	0.859	0.467	0.663	0.633

FOLD 4						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.689	0.874	0.368	0.621	0.567
0.1	1	0.709	0.867	0.435	0.651	0.614
0.1	10	0.712	0.864	0.447	0.656	0.621
0.1	100	0.712	0.864	0.447	0.656	0.621
1	0.1	0.697	0.857	0.42	0.639	0.600
1	1	0.715	0.855	0.473	0.664	0.636
1	10	0.716	0.853	0.477	0.665	0.638
1	100	0.716	0.853	0.477	0.665	0.638
10	0.1	0.702	0.848	0.451	0.650	0.618
10	1	0.716	0.853	0.477	0.665	0.638
10	10	0.716	0.853	0.477	0.665	0.638
10	100	0.716	0.853	0.477	0.665	0.638
100	0.1	0.707	0.855	0.45	0.653	0.620
100	1	0.716	0.853	0.477	0.665	0.638
100	10	0.716	0.853	0.477	0.665	0.638
100	100	0.716	0.853	0.477	0.665	0.638

FOLD 5						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.691	0.873	0.371	0.622	0.569
0.1	1	0.71	0.871	0.43	0.651	0.612
0.1	10	0.712	0.874	0.429	0.652	0.612
0.1	100	0.712	0.874	0.429	0.652	0.612
1	0.1	0.698	0.862	0.413	0.638	0.597
1	1	0.716	0.854	0.476	0.665	0.638
1	10	0.716	0.853	0.478	0.666	0.639
1	100	0.716	0.853	0.478	0.666	0.639
10	0.1	0.703	0.852	0.444	0.648	0.615
10	1	0.716	0.853	0.478	0.666	0.639
10	10	0.716	0.853	0.478	0.666	0.639
10	100	0.716	0.853	0.478	0.666	0.639
100	0.1	0.708	0.849	0.463	0.656	0.627
100	1	0.716	0.853	0.478	0.666	0.639
100	10	0.716	0.853	0.478	0.666	0.639
100	100	0.716	0.853	0.478	0.666	0.639

FOLD 6						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.689	0.875	0.364	0.620	0.564
0.1	1	0.71	0.869	0.434	0.652	0.614
0.1	10	0.712	0.871	0.435	0.653	0.616
0.1	100	0.712	0.871	0.435	0.653	0.616
1	0.1	0.697	0.854	0.424	0.639	0.602
1	1	0.716	0.843	0.494	0.669	0.645
1	10	0.716	0.844	0.494	0.669	0.646
1	100	0.716	0.844	0.494	0.669	0.646
10	0.1	0.704	0.847	0.454	0.651	0.620
10	1	0.716	0.843	0.495	0.669	0.646
10	10	0.716	0.844	0.494	0.669	0.646
10	100	0.716	0.844	0.494	0.669	0.646
100	0.1	0.707	0.853	0.455	0.654	0.623
100	1	0.716	0.843	0.495	0.669	0.646
100	10	0.716	0.844	0.494	0.669	0.646
100	100	0.716	0.844	0.494	0.669	0.646

FOLD 7						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.691	0.87	0.379	0.625	0.574
0.1	1	0.709	0.871	0.428	0.650	0.611
0.1	10	0.711	0.868	0.439	0.654	0.617
0.1	100	0.711	0.868	0.439	0.654	0.617
1	0.1	0.698	0.859	0.416	0.638	0.598
1	1	0.716	0.853	0.477	0.665	0.638
1	10	0.716	0.853	0.477	0.665	0.638
1	100	0.716	0.853	0.477	0.665	0.638
10	0.1	0.703	0.855	0.438	0.647	0.612
10	1	0.716	0.852	0.478	0.665	0.638
10	10	0.716	0.853	0.477	0.665	0.638
10	100	0.716	0.853	0.477	0.665	0.638
100	0.1	0.708	0.849	0.462	0.656	0.626
100	1	0.716	0.852	0.478	0.665	0.638
100	10	0.716	0.853	0.477	0.665	0.638
100	100	0.716	0.853	0.477	0.665	0.638

FOLD 8						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.691	0.871	0.378	0.625	0.574
0.1	1	0.71	0.868	0.437	0.653	0.616
0.1	10	0.711	0.87	0.435	0.653	0.615
0.1	100	0.711	0.87	0.435	0.653	0.615
1	0.1	0.698	0.859	0.416	0.638	0.598
1	1	0.716	0.857	0.47	0.664	0.635
1	10	0.716	0.857	0.472	0.665	0.636
1	100	0.716	0.857	0.472	0.665	0.636
10	0.1	0.704	0.853	0.446	0.650	0.617
10	1	0.716	0.852	0.478	0.665	0.638
10	10	0.716	0.857	0.472	0.665	0.636
10	100	0.716	0.857	0.472	0.665	0.636
100	0.1	0.716	0.857	0.472	0.665	0.636
100	1	0.716	0.852	0.478	0.665	0.638
100	10	0.716	0.857	0.472	0.665	0.636
100	100	0.716	0.857	0.472	0.665	0.636

FOLD 9						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.689	0.871	0.372	0.622	0.569
0.1	1	0.71	0.866	0.436	0.651	0.614
0.1	10	0.711	0.868	0.437	0.653	0.616
0.1	100	0.711	0.868	0.437	0.653	0.616
1	0.1	0.697	0.854	0.424	0.639	0.602
1	1	0.715	0.847	0.485	0.666	0.641
1	10	0.715	0.847	0.486	0.667	0.642
1	100	0.715	0.847	0.486	0.667	0.642
10	0.1	0.703	0.855	0.437	0.646	0.611
10	1	0.715	0.845	0.489	0.667	0.643
10	10	0.715	0.847	0.486	0.667	0.642
10	100	0.715	0.847	0.486	0.667	0.642
100	0.1	0.715	0.847	0.486	0.667	0.642
100	1	0.715	0.845	0.489	0.667	0.643
100	10	0.715	0.847	0.486	0.667	0.642
100	100	0.715	0.847	0.486	0.667	0.642

FOLD 10						
COST	GAMMA	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.1	0.1	0.689	0.874	0.368	0.621	0.567
0.1	1	0.71	0.873	0.426	0.650	0.610
0.1	10	0.711	0.872	0.431	0.652	0.613
0.1	100	0.711	0.872	0.431	0.652	0.613
1	0.1	0.698	0.861	0.414	0.638	0.597
1	1	0.715	0.856	0.47	0.663	0.634
1	10	0.715	0.851	0.48	0.666	0.639
1	100	0.715	0.851	0.48	0.666	0.639
10	0.1	0.703	0.854	0.442	0.648	0.614
10	1	0.715	0.85	0.48	0.665	0.639
10	10	0.715	0.851	0.48	0.666	0.639
10	100	0.715	0.851	0.48	0.666	0.639
100	0.1	0.707	0.853	0.455	0.654	0.623
100	1	0.715	0.85	0.48	0.665	0.639
100	10	0.715	0.851	0.48	0.666	0.639
100	100	0.715	0.851	0.48	0.666	0.639



### Lampiran 5. Output *Support Vector Machine* Kernel Linier

FOLD 1					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.689	0.848	0.41	0.629	0.590
0.1	0.69	0.846	0.419	0.633	0.595
1	0.69	0.846	0.419	0.633	0.595
10	0.69	0.846	0.417	0.632	0.594
100	0.685	0.833	0.42	0.627	0.591

FOLD 2					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.689	0.847	0.416	0.632	0.594
0.1	0.689	0.845	0.418	0.632	0.594
1	0.689	0.845	0.418	0.632	0.594
10	0.689	0.845	0.418	0.632	0.594
100	0.686	0.843	0.413	0.628	0.590

FOLD 3					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.689	0.855	0.399	0.627	0.584
0.1	0.689	0.854	0.402	0.628	0.586
1	0.689	0.853	0.403	0.628	0.586
10	0.688	0.852	0.404	0.628	0.587
100	0.686	0.84	0.416	0.628	0.591

FOLD 4					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.688	0.845	0.418	0.632	0.594
0.1	0.689	0.843	0.424	0.634	0.598
1	0.689	0.843	0.424	0.634	0.598
10	0.689	0.843	0.424	0.634	0.598
100	0.686	0.835	0.425	0.630	0.596

FOLD 5					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.689	0.846	0.413	0.630	0.591
0.1	0.689	0.844	0.42	0.632	0.595
1	0.689	0.844	0.42	0.632	0.595
10	0.689	0.845	0.42	0.633	0.596
100	0.686	0.838	0.422	0.630	0.595

FOLD 6					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.69	0.843	0.425	0.634	0.599
0.1	0.69	0.842	0.426	0.634	0.599
1	0.69	0.841	0.427	0.634	0.599
10	0.691	0.842	0.426	0.634	0.599
100	0.691	0.842	0.426	0.634	0.599

FOLD 7					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.69	0.849	0.415	0.632	0.594
0.1	0.691	0.848	0.417	0.633	0.595
1	0.691	0.848	0.417	0.633	0.595
10	0.691	0.848	0.416	0.632	0.594
100	0.689	0.84	0.426	0.633	0.598

FOLD 8					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.69	0.841	0.43	0.636	0.601
0.1	0.69	0.841	0.43	0.636	0.601
1	0.69	0.841	0.43	0.636	0.601
10	0.69	0.835	0.437	0.636	0.604
100	0.69	0.835	0.437	0.636	0.604

FOLD 9					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.69	0.84	0.43	0.635	0.601
0.1	0.69	0.839	0.43	0.635	0.601
1	0.69	0.839	0.43	0.635	0.601
10	0.689	0.839	0.431	0.635	0.601
100	0.689	0.839	0.431	0.635	0.601

FOLD 10					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.692	0.845	0.424	0.635	0.599
0.1	0.691	0.844	0.426	0.635	0.600
1	0.692	0.844	0.426	0.635	0.600
10	0.69	0.854	0.405	0.630	0.588
100	0.699	0.856	0.398	0.627	0.584

## Lampiran 6. Output *Support Vector Machine* Kernel Linier Seleksi Variabel

FOLD 1					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.687	0.835	0.431	0.633	0.600
0.1	0.687	0.835	0.431	0.633	0.600
1	0.687	0.835	0.431	0.633	0.600
10	0.687	0.835	0.431	0.633	0.600
100	0.687	0.835	0.431	0.633	0.600

FOLD 2					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.687	0.834	0.432	0.633	0.600
0.1	0.687	0.834	0.432	0.633	0.600
1	0.687	0.834	0.432	0.633	0.600
10	0.687	0.834	0.432	0.633	0.600
100	0.684	0.829	0.435	0.632	0.601

FOLD 3					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.687	0.835	0.431	0.633	0.600
0.1	0.687	0.835	0.431	0.633	0.600
1	0.687	0.835	0.431	0.633	0.600
10	0.687	0.835	0.431	0.633	0.600
100	0.687	0.835	0.431	0.633	0.600

FOLD 4					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.686	0.835	0.429	0.632	0.599
0.1	0.686	0.835	0.429	0.632	0.599
1	0.686	0.835	0.429	0.632	0.599
10	0.686	0.835	0.429	0.632	0.599
100	0.686	0.835	0.429	0.632	0.599

FOLD 5					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.688	0.835	0.432	0.634	0.601
0.1	0.688	0.835	0.432	0.634	0.601
1	0.688	0.835	0.432	0.634	0.601
10	0.688	0.835	0.432	0.634	0.601
100	0.688	0.835	0.432	0.634	0.601

FOLD 6					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.687	0.833	0.432	0.633	0.600
0.1	0.687	0.833	0.432	0.633	0.600
1	0.687	0.833	0.432	0.633	0.600
10	0.687	0.833	0.432	0.633	0.600
100	0.687	0.833	0.432	0.633	0.600

FOLD 7					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.688	0.834	0.434	0.634	0.602
0.1	0.688	0.834	0.434	0.634	0.602
1	0.688	0.834	0.434	0.634	0.602
10	0.688	0.834	0.434	0.634	0.602
100	0.688	0.834	0.434	0.634	0.602

FOLD 8					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.687	0.834	0.433	0.634	0.601
0.1	0.687	0.834	0.433	0.634	0.601
1	0.687	0.834	0.433	0.634	0.601
10	0.687	0.834	0.433	0.634	0.601
100	0.687	0.834	0.433	0.634	0.601

FOLD 9					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.687	0.833	0.433	0.633	0.601
0.1	0.687	0.833	0.433	0.633	0.601
1	0.687	0.833	0.433	0.633	0.601
10	0.687	0.833	0.433	0.633	0.601
100	0.682	0.823	0.438	0.631	0.600

FOLD 10					
COST	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
0.01	0.687	0.834	0.433	0.634	0.601
0.1	0.687	0.834	0.433	0.634	0.601
1	0.687	0.834	0.433	0.634	0.601
10	0.687	0.834	0.433	0.634	0.601
100	0.687	0.834	0.433	0.634	0.601

### Lampiran 7. Output *Naïve Bayes Classifier*

TANPA SELEKSI VARIABEL					
Fold	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
1	0.669	0.736	0.553	0.645	0.638
2	0.679	0.746	0.56	0.653	0.646
3	0.678	0.742	0.567	0.655	0.649
4	0.679	0.735	0.576	0.656	0.651
5	0.668	0.731	0.557	0.644	0.638
6	0.673	0.746	0.546	0.646	0.638
7	0.67	0.744	0.545	0.645	0.637
8	0.666	0.741	0.534	0.638	0.629
9	0.675	0.744	0.554	0.649	0.642
10	0.673	0.741	0.557	0.649	0.642

SELEKSI VARIABEL FCBF					
Fold	AKURASI	RECALL 1	RECALL 2	AUC	G_MEAN
1	0.688	0.743	0.539	0.641	0.633
2	0.685	0.761	0.547	0.654	0.645
3	0.676	0.745	0.557	0.651	0.644
4	0.673	0.741	0.552	0.647	0.640
5	0.671	0.741	0.551	0.646	0.639
6	0.676	0.757	0.535	0.646	0.636
7	0.67	0.753	0.528	0.641	0.631
8	0.666	0.748	0.521	0.635	0.624
9	0.677	0.756	0.536	0.646	0.637
10	0.675	0.753	0.54	0.647	0.638

## **BIODATA PENULIS**



Penulis yang bernama lengkap Wiwin Yuli Widiawati lahir di Pamekasan, 29 Juli 1994 sebagai anak sulung dari dua bersaudara. Pendidikan formal yang telah ditempuh berawal dari SDN Pademawu Timur III, SMPN 1 Pademawu, SMAN 1 Pamekasan. Pada tahun 2013 penulis diterima sebagai Mahasiswa DIII Statistika

ITS dengan NRP 1313030074 dan pada tahun 2016 penulis diterima sebagai Mahasiswa S1 Lintas Jalur Departemen Statistika ITS dengan NRP 06211645000015. Selama masa perkuliahan, di samping kegiatan akademik, penulis juga terlibat aktif dalam kegiatan kemahasiswaan. Penulis pernah sebagai staf Departemen Dalam Negeri (DAGRI) HIMADATA ITS. Pada masa perkuliahan, penulis pernah melakukan kerja praktek di Kantor Wilayah Bea dan Cukai Malang, khususnya pada Departemen Penindakan dan Penyidikan. Pembaca yang ingin berdiskusi, memberi kritik dan saran dapat melalui [wiwinyuliwidiawati@gmail.com](mailto:wiwinyuliwidiawati@gmail.com).



## SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS:

Nama : Wiwin Yuli Widiawati

NRP : 06211645000015

menyatakan bahwa data yang digunakan dalam Tugas Akhir/ Thesis ini merupakan data sekunder yang diambil dari ~~penelitian/buku/Tugas Akhir/Thesis~~ publikasi lainnya yaitu:

Sumber : PT PLN (Persero) Distribusi Jawa Timur

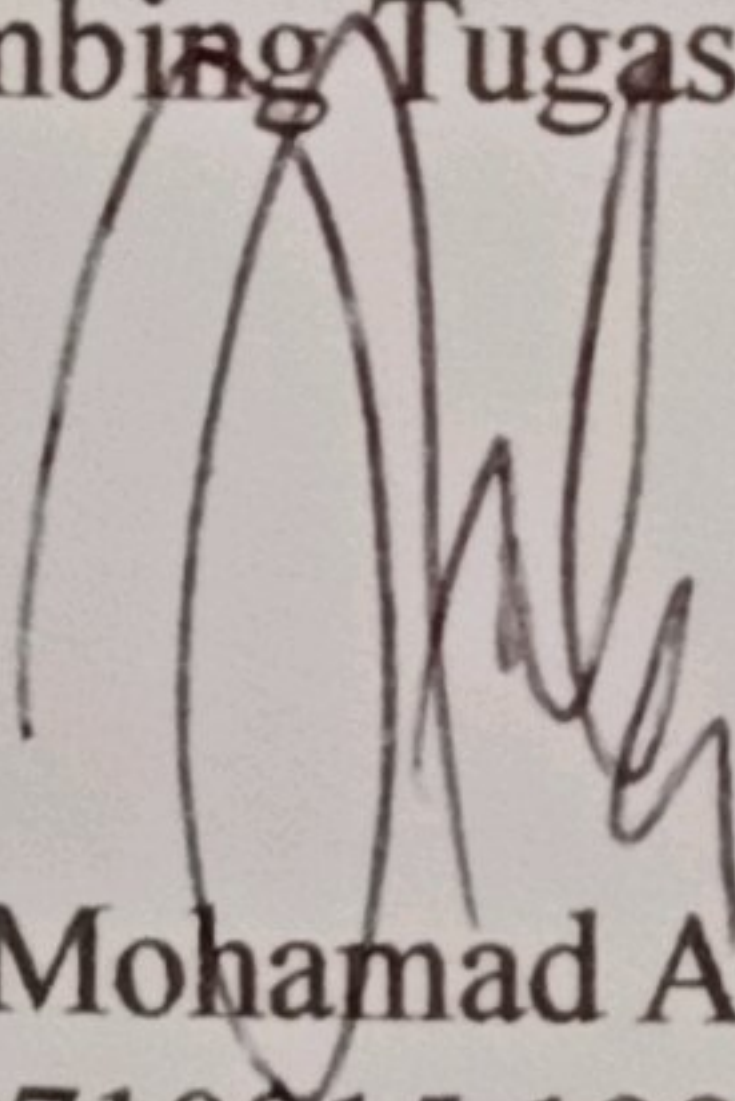
Keterangan :

Data pelanggan rumah tangga yang memperoleh subsidi listrik bulan Januari 2018 yaitu pelanggan rumah tangga dengan daya listrik 450 VA dan 900 VA.

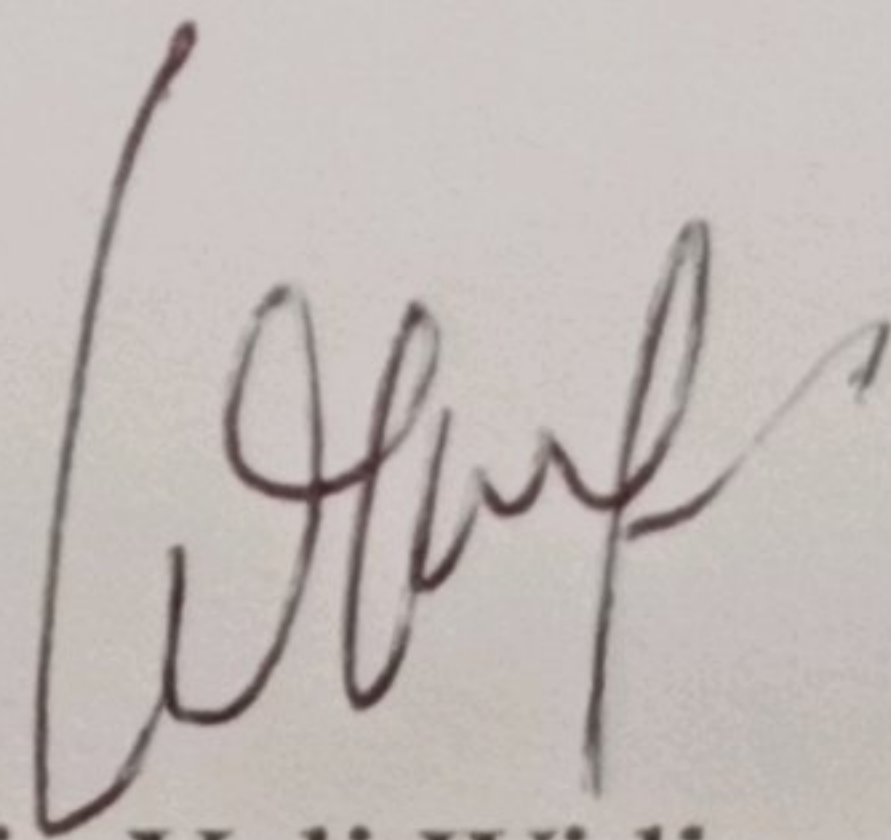
Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Surabaya, Juli 2018

Mengetahui,  
Pembimbing Tugas Akhir



Raden Mohamad Atok, Ph.D  
NIP. 19710915 199702 1 001



Wiwin Yuli Widiawati  
NRP. 06211645000015